A GENE PAIR FROM THE JUNGLE

Discovery, Characterization and Applications of the *Enterobacter lignolyticus* Eil-Efflux system

Inauguraldissertation

zur
Erlangung der Würde eines Doktors der Philosophie
vorgelegt der
Philosophisch-Naturwissenschaftlichen Fakultät
der Universität Basel
von

Thomas Lawrence Rüegg

aus Muttenz, BL

Basel, 2017

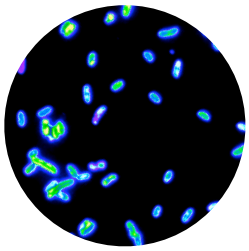
Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät auf Antrag von Professor Dr. Thomas Boller und Professor Dr. Martin Ackermann.

Basel, den 23. Juni 2015

Prof. Dr. J. Schibler, Dekan

A Gene Pair from the Jungle

Discovery, Characterization and Applications of the *Enterobacter lignolyticus* EilEfflux System



Inauguraldissertation

zur Erlangung der Würde eines Doktors der Philosophie
vorgelegt der Philosophisch-Naturwissenschaftlichen Fakultät der
Universität Basel

Von Thomas Lawrence Rüegg aus Muttenz, Baselland
Basel, 2015

Table of Contents

SUMMARY OF THIS THESIS

The findings presented in this work emerge from the discovery of a multidrug efflux system *Enterobacter lignolyticus*, a bacterium isolated from Puerto Rican cloud forest soil. This system, which consists of the inner membrane transporter EilA and its cognate repressor EilR, confers tolerance to imidazolium-based ionic liquids. In my research, I characterized these genes using molecular genetics, improved a method of microbial biofuel production using synthetic biology techniques, and demonstrated new biotechnological applications by both decoupling the genes from their natural context and engineering the regulatory elements.

Chapter 1 targets the removal of a bottleneck in the microbial conversion of lignocellulose to biofuels or chemicals. In this process, pretreatment of plant biomass is necessary due to the inherent recalcitrance of lignocellulose. Certain ionic liquids (ILs) are solvents remarkably effective in solubilizing cellulosic biomass. By dissociating lignin from hemicellulose and cellulose in cell walls, enzymatic hydrolysis to fermentable sugars can be achieved. However, ILs are toxic to most microbes, inhibiting growth and subsequent fermentation of sugars to fuels. In searching for a solution to this problem, I discovered a novel molecular system for bacterial resistance to IL toxicity by screening the genome of the IL-tolerant bacterium *E. lignolyticus*. A single gene was identified that promotes growth in the presence of IL, namely, an multidrug transporter, EilA, which acts to export IL from the cell. In response to changes in external IL levels, expression of the transporter is controlled by a repressor, EilR, providing a self regulating system that maintains cell viability. The gene pair encoding EilA and EilR remains functional when transferred to an *Escherichia coli* strain that expresses a biosynthetic pathway to produce a terpene-based biofuel. In this host organism, the auto-regulatory efflux system enables growth and biofuel production in a previously toxic environment.

Chapter 2 focuses on the EilR protein and how it performs its task as a transcriptional regulator. Identification of the DNA binding site, the eil-operator, provided the basis to develop a sensitive EilR-regulated promoter that drives expression of a reporter gene in *E. coli.* Using this cellular biosensor, I identified a range of cationic dyes with high affinity to the EilR repressor. These anthropogenic ligands are unrelated to ILs, and some of them can induce the biosensor at

nanomolar concentrations – up to five orders of magnitude lower than ILs. Activity-based assays with cell extracts indicated a possible way of identifying metabolites as natural inducers. In addition, experiments with homologous repressors and a transporter provided further insights on bacterial regulation of efflux.

The strong binding affinity of EilR to its operator and to the readily available ligands motivated me to use this mechanism to develop an inducible system for gene expression (**Chapter 3**). The three approaches taken to achieve this goal comprise 1) targeted modifications in the native promoter region, 2) refinements of the biosensor promoter and, 3) insertion of operator sites into early promoters from bacteriophages. Using this approach, I generated a set of tightly repressible promoters that are – upon addition of the characterized effector molecules – inducible over more than four orders of magnitude, reaching expression levels comparable to those of the strongest characterized expression systems. Besides *Escherichia coli,* these promoters are functional in other distantly related bacteria, such as *Pseudomonas putida* and *Sinorhizobium meliloti.* I then introduced the EilR-guided regulatory mechanism into the yeast *Saccharomyces cerevisiae* to show how this bacterial repressor can control the activity of a modified yeast promoter containing EilR-binding sites. The EilR-based molecular switch can therefore serve as a tool for independent gene regulation in prokaryotic and eukaryotic organisms.

General Introduction

I find it fascinating to explore natural processes that provide a beneficial feedback in the broadest sense – the collection or cultivation of mushrooms, catching a trout at a lonely mountain lake, collecting tasty fruits in the rainforest, sniffing at a fragrant flower, or conducting research to discover biological functions that can be translated into applications.

The consumption of fossil fuels during a period of intensive traveling contributed to my motivation to combine my scientific curiosity with an aspiration of gaining an insight into and contributing to our understanding of the field of biofuels. Thus, I chose to pursue my doctoral research in an environment that mainly focuses on the microbial conversion of lignocellulosic biomass into fuels and chemicals.

I started my work by entering the dense jungle of genetic information stored in environmental microorganisms with the primary aim to improve microbial performance in growth during biofuel production. During the course of my experiments, the results inspired me to define an additional objective to develop generally applicable tool for controlling gene expression, which could also be useful in biofuel production.

The foundation of my research is built on the initial discovery of two associated genes that encode a bacterial efflux system for “ionic liquids”, organic salts that are important solvents used in biofuel production (as described below). In Chapter 1, I describe the approach that led to the discovery, as well as the relevance of this efflux system in biofuel applications. Chapter 2 focuses on the characterization of the regulatory aspects of the efflux system. This chapter also builds the transition from my original objective to the emergence of a new, application-oriented goal, which I pursue in the next chapter. In Chapter 3, I decouple the regulatory elements of the efflux system from its native context and use these elements to develop an inducible gene expression system. The three chapters form the body of this thesis and are mostly autonomous, each providing elaborate discussions and conclusions. Towards the end of this thesis, I summarize my main conclusions in a brief final statement.

In the following sections of this general introduction, I delineate the background of my research projects.

Ionic liquids as agents for biofuel production

The major part of terrestrial plant biomass consists of the three cell wall polymers, cellulose, hemicellulose and lignin. Thus, lignocellulosic biomass is an abundant resource available for the sustainable production of biofuels and high value chemicals. Some of the most promising approaches have centered on engineering microbes (Liu & Khosla, 2010; Rabinovitch-Deere *et al.*, 2013; Wen *et al.*, 2013) and utilizing a wide range of feedstocks, including woody biomass, indigenous grasses, and agricultural residues such as corn stover (Bokinsky *et al.*, 2011; de Jong *et al.*, 2012).

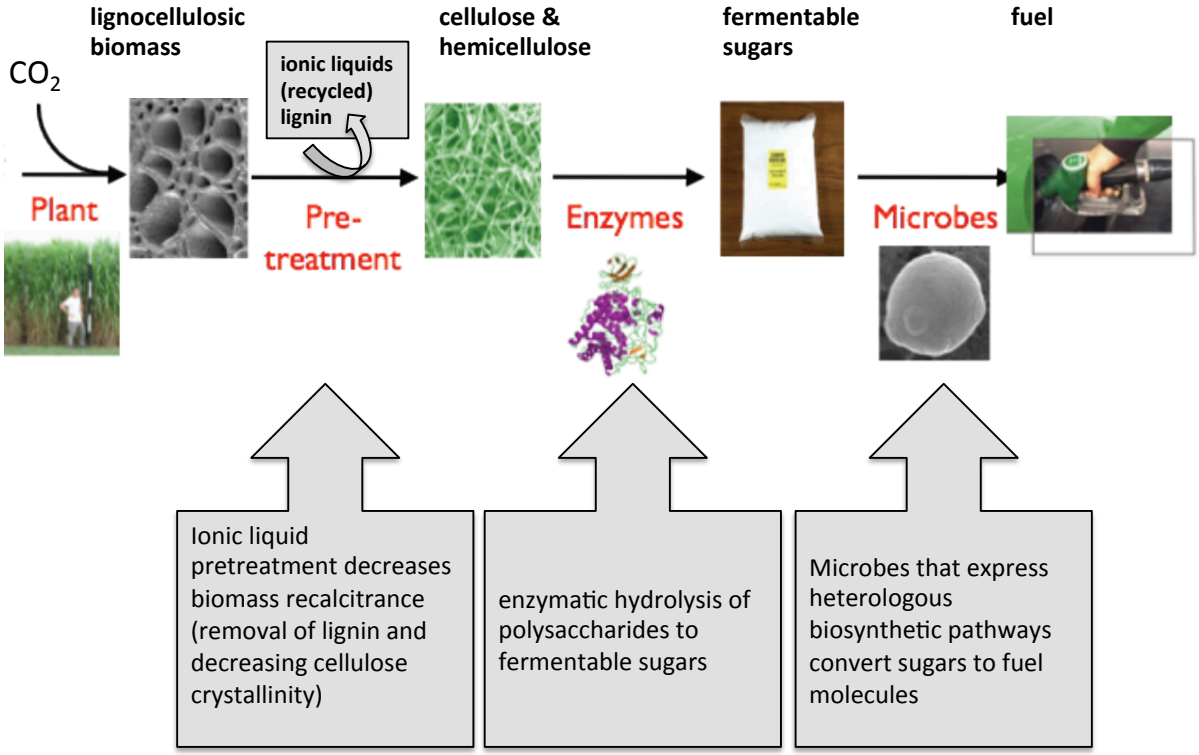


Figure 1 | Approach taken at the Joint BioEnergy Institute to convert biomass to fuels. Lignocellulosic biomass is pretreated by recyclable ILs in order to decrease biomass recalcitrance, mainly by removing lignin and decreasing cellulose crystallinity. This procedure facilitates the subsequent enzymatic digestion of the cell wall polysaccharides to fermentable sugars. Engineered microorganisms utilize these sugars to produce fuel molecules via the heterologous expression of biosynthetic pathways (adopted from JBEI.org).

Lignocellulose is an extremely recalcitrant material, in which cellulose forms highly crystalline micro-fibrils that are embedded in a matrix consisting of hemicellulose and lignin. By linking to hemicellulose and cellulose, lignin acts as a barrier that prevents enzymes from gaining access to and hydrolyzing the polysaccharides (Dashtban *et al.*, 2010).

Pretreatment of lignocellulosic biomass is a crucial step, since it removes lignin from the polysaccharides and reduces the crystallinity of cellulose, which is required for an efficient subsequent enzymatic hydrolysis of polysaccharides. Lignocellulosic biomass is commonly pretreated with dilute acid (Grohmann *et al.*, 1986), ammonia (Dale & Moreira, 1982), steam (Bobleter *et al.*, 1976) and, more recently, ionic liquids.

Ionic liquids (ILs) are salts that are composed of organic cations and either organic or inorganic anions. Different cation-anion combinations give rise to an almost unlimited number of ILs (Brandt *et al.*, 2013; Keskin *et al.*, 2007), some of which have emerged as remarkably efficient agents for the pretreatment of lignocellulose. The most common ILs can be divided into four major groups that are defined by their cations: Quaternary ammonium ILs, N-alkylpyridinium ILs, N-alkyl-isoquinolinium ILs, and 1-alkyl-3-methylimidazolium ILs (Figure 1). Among them, imidazolium-based cations form the most widely used ILs for biomass pretreatment (Brandt *et al.*, 2013). They effectively solubilize lignocellulose by disrupting inter- and intra-molecular hydrogen bonds, enabling enzymatic hydrolysis to generate high yields of fermentable sugars, exceeding those resulting from conventional pretreatment methods, (C. Li et al., 2010, 2011). In addition, ILs generate comparatively low amounts of biomass-derived side products, such as aromatic lignin monomers or polysaccharide-derived aldehydes, which inhibit downstream fermentation (Blanch *et al.*, 2011; C. L. Li *et al.*, 2010; Mora-Pale *et al.*, 2011).

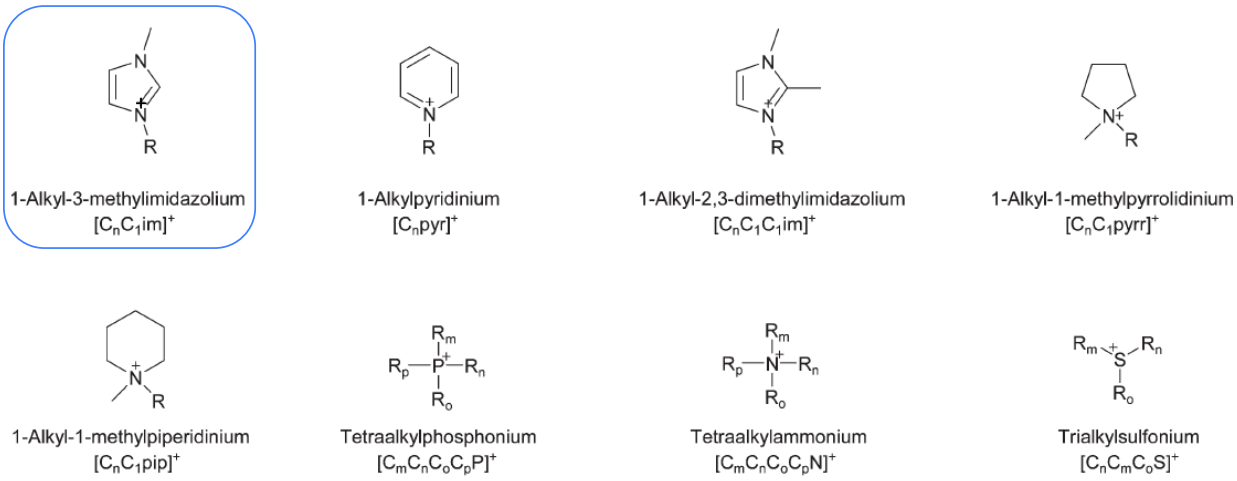


Figure 2 | Common ionic liquid cations. The focus in this work is on 1-alkyl-3-methylimidazolium cations (framed in blue), namely $[C_2C_1im]^+$ (also written as $[C_2mim]^+$), which contains an ethyl-group for R. Figure adopted from Brandt *et al.*, 2013**.**

A drawback of these ILs is their toxicity to fermentative microbes, such as *Escherichia coli* (Lee & Chang, 2005; Park et al., 2012; Ranke et al., 2007) and *Saccharomyces cerevisiae* (Ouellet *et al.*, 2011), as well as their inhibition of hydrolyzing enzymes. For these reasons, and since the extensive washing required for complete IL removal is not feasible in large-scale applications, residual IL present in the pretreated biomass lowers efficiency of microbial biofuel production. An ideal and more sustainable process should balance the costs of removing IL with fermentation performance (Klein-Marcuschamer *et al.*, 2011). A novel way to achieve this would use biofuel producing microbes that can tolerate residual levels (e.g. 0.2-5% wt/vol) of ILs.

Microbial response to ionic liquids

Toxicity of ILs to microbes is due to their detergent-like effect causing membrane disruption and subsequent accumulation within cells (Łuczak *et al.*, 2010). ILs possessing long hydrophobic alkyl side chains additionally promote toxicity by increasing their membrane interactions, thereby making them candidate antimicrobial agents (Łuczak *et al.*, 2010; Ranke *et al.*, 2007; Samorì, 2011).

A few environmental bacteria (DeAngelis *et al.*, 2011; Megaw *et al.*, 2013) and fungi (Huang *et al.*, 2013; Sitepu *et al.*, 2014) have been reported to tolerate ILs (Huang *et al.*, 2013; Khudyakov *et al.*, 2012; Megaw *et al.*, 2013; Sitepu *et al.*, 2014), but a mechanism for microbial tolerance to ILs has not been elucidated. Because these microbes should possess genes responsible for IL tolerance, I began investigating one such bacterium, *Enterobacter lignolyticus* SCF1, a facultative anaerobic gamma-proteobacterium isolated from tropical cloud forest soil in the Puerto Rican El Yunque Forest. While this bacterium was detected during an attempt to discover bacteria capable of anaerobic degradation of lignin, it was also revealed that this organism grew well in cultures containing 380 mM (5.5%) of 1-ethyl-3-methylimidazolium chloride[1] (DeAngelis *et al.*, 2011; Khudyakov *et al.*, 2012). This IL (abbreviated as [C_2mim]Cl, see), has been identified as a promising pretreatment solvent for biomass (Mora-Pale *et al.*, 2011). Using RNA-sequencing transcriptomics, a former group member then found that *E. lignolyticus* responds to [C_2mim]Cl by the differential expression of several hundred genes, resulting in a complex response pattern with numerous phenotypical changes (Khudyakov *et al.*, 2012). These include the synthesis and import of osmoprotectants, upregulation of membrane

[1] I define IL tolerance here and throughout as normal or near normal bacterial growth in media containing up to 270 mM (~4 % [w/v]), and some depression in growth rate and yield between 270 – 415 mM (~2 – 6 % [w/v]) imidazolium cation. These concentrations are given in the context of the fermentation stage of biofuel production when IL is used as the pre-treatment and then contaminates the downstream process.

transporters, and a decrease in membrane permeability due to the downregulation of porins and the production of cyclopropane fatty acids as bulky membrane lipids. Among the induction of transcripts, the upregulation of genes encoding putative efflux pumps, which includes multidrug transporters, is most relevant to my studies.

Multidrug transporters

Active membrane transporters comprise a relatively small number of protein superfamilies that either require ATP or make use of the electrochemical gradient along the cell membrane, acting either as proton symporter, uniporter or antiporter (Henderson & Maiden, 1990; Pao *et al.*, 1998). Such transporters are typically highly specific for their substrate, such as a sugar, an amino acid, or a peptide. In contrast, multidrug transporters have broad specificity for a wide range of chemically dissimilar molecules. Such multidrug resistance (MDR) pumps are represented in five superfamilies (Krulwich *et al.*, 2005), including the major facilitator superfamily (MFS), to which the EilA pump described in Chapter 1 belongs. Substrates of multidrug transporters are usually relatively hydrophobic, planar molecules of a molecular weight less than 800 Da, which are in most cases weakly cationic (Higgins, 2007).

Occurrence of multidrug transporters in several phylogenetically distant superfamilies indicates that the ability to export drugs is not a recent response to antimicrobial chemotherapy. Instead, it is rather a stably conserved trait that emerged on only a few evolutionary events (Saier & Paulsen, 2001). The similar abundance of chromosomally encoded multidrug efflux pumps both in pathogenic and nonpathogenic bacteria (Saier *et al.*, 1998) indicates that these proteins already fulfilled certain physiological functions before the MDR phenotype evolved (Krulwich *et al.*, 2005). The corresponding genes could undergo mutations to produce fully functional MDR proteins without significant loss of the original function (Krulwich *et al.*, 2005). As such, although MDR-transporters often have overlapping substrate preferences, they can still pursue additional differential, more specific physiological roles. For example overexpression of both *Bacillus subtilis* Blt and *Escherichia coli* MdfA transporters confer tolerance to ethidium bromide in their native hosts. At the same time, physiological levels of Blt contribute to spermidine metabolism (Ahmed *et al.*, 1995), while MdfA is involved in pH homeostasis as a Na^+/H^+ antiporter (Lewinson *et al.*, 2004). Such observations indicate that the broad specificity towards inhibitory compounds is actually rather an opportunistic effect of a more specific physiological function rather than an evolutionary driver. This hypothesis is additionally supported by a certain functional redundancy of MDR-transporters within

the same organism. In *Bacillus subtilis* for example, the Blt and Bmr transporters have common drug substrates. However, their cognate transcriptional regulators influence the resistance phenotype by responding differently to these drugs (Ahmed *et al.*, 1995).

Heterologous gene expression in biotechnology

Biotechnological applications utilize the ability of microorganisms to express heterologous genes and gene pathways for the production of proteins and small molecules. Among the many available microorganisms, the bacterium *E. coli* and the yeast *S. cerevisiae* are the best characterized and most commonly used organisms. In biotechnology, a portion of the available nutrient, including lignocellulose-derived fermentable sugars (Bokinsky *et al.*, 2011; Steen *et al.*, 2010), is deviated towards to the pathways that synthesize the desired product. The expression of heterologous biosynthetic pathways opens the way for the microbial production of a wide variety of compounds, including alkaloids (Brown *et al.*, 2015), polyketides (Pfeifer *et al.*, 2001) and terpenes (Martin *et al.*, 2003). For example, the engineered mevalonate pathway in conjunction with terpene synthases can serve as platform to produce a wide range of terpenes, such as precursors for anticancer or antimalarial drugs (Ajikumar *et al.*, 2010; Ro *et al.*, 2006) or candidate biofuels (Peralta-Yahya *et al.*, 2011) (Figure 3).

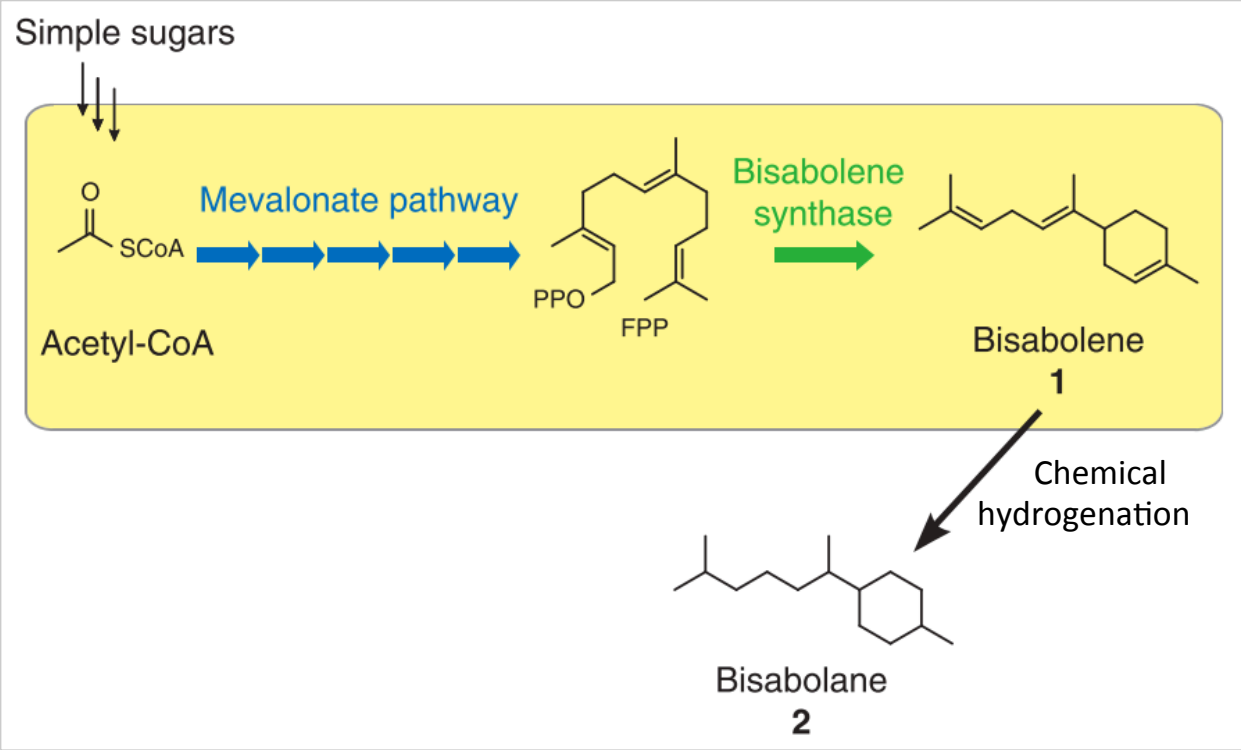


Figure 3 | Example of a pathway for the biosynthesis of bisabolene, a terpeneoid biofuel precursor. The engineered microbe (yellow box) converts simple sugars (e.g. from biomass hydrolysate) into acetyl-CoA via primary metabolism. The heterologous mevalonate pathway

converts acetyl-CoA into farnesyl pyrophosphate (FPP), which is then converted into bisabolene by a plant-derived terpene synthase. Chemical hydrogenation of biosynthetic bisabolene leads to bisabolane, an alternative to fossil diesel (Adopted from Peralta-Yahya *et al.*, 2011).

The ability to control gene transcription is essential for the fundamental understanding of gene function and for regulating expression levels in biotechnological applications.

One of the major challenges in biotechnology is that heterologously expressed gene products are often toxic and overexpression causes a general metabolic burden on the cells, resulting in poor growth rates and low yields (Maya *et al.*, 2008). Inducible promoters are ideal for adjusting expression levels and for decoupling growth from production, which enables the establishment of a healthy culture prior to experimentally directed expression. Control over gene expression by such promoters is in most cases enabled by a transcription factor that specifically responds to an externally added small molecule.

These transcription factors specifically bind to their DNA binding sites (operators) located in the promoter region. Upon binding with the signal molecule, the transcription factor undergoes a conformational change, which alters binding affinity to its operator, resulting in gene transcription. Such molecular interactions have been adopted and altered in many ways to control expression of heterologous target genes or gene pathways.

Inducible gene expression systems in *E. coli*

The most common bacterial induction systems, some of which are listed in Table 1, make use of transcription factors that originally regulate functional genes as a response to environmental signals, such as a carbon source or a toxin. For example, the lactose-responsive LacI repressor regulates the lactose utilization pathway via the lac-promoter (Kennell & Riezman, 1977); the arabinose-responsive AraC activator guides expression the arabinose pathway (Guzman *et al.*, 1995), while tetracycline triggers the expression of a tetracycline efflux pump via the TetR repressor (Hillen *et al.*, 1983). Tet-promoters that are regulated by TetR are commonly used when tight regulation is required (Skerra, 1994). The AraC regulated pBAD-promoters provide the most tightly regulated control over gene expression

(Guzman *et al.*, 1995; Terpe, 2006), which is especially important when basal gene expression in the non-induced state causes cellular stress.

Table 1 | Some *E. coli* promoter systems that are in use for heterologous protein production and their characteristics (adopted from Terpe, 2006)

Expression system based on	Induction (range of inductor)	Level of expression	Key features
lac promoter	Addition of IPTG 0.2 mM (0.05–2.0 mM)	Low level up to middle	Weak, regulated suitable for gene products at very low intracellular level Comparatively expensive induction
trc and *tac* promoter	Addition of IPTG 0.2 mM (0.05–2.0 mM)	Moderately high	High-level, but lower than T7 system Regulated expression still possible Comparatively expensive induction High basal level
T7 RNA polymerase	Addition of IPTG 0.2 mM (0.05–2.0 mM)	Very high	Utilizes T7 RNA polymerase High-level inducible over expression T7*lac* system for tight control of induction needed for more toxic clones Relative expensive induction Basal level depends on used strain (pLys)
Phage promoter p_L	Shifting the temperature from 30 to 42 °C (45 °C)	Moderately high	Temperature-sensitive host required Less likelihood of "leaky" uninduced expression Basal level, high basal level by temperatures below 30 °C No inducer
tetA promoter/operator	Anhydrotetracycline 200μg/l	Variable from middle to high level	Tight regulation Independent on metabolic state Independent on *E. coli* strain Relative inexpensive inducer Low basal level
araBAD promoter (P_{BAD})	Addition of L-arabinose 0.2 % (0.001–1.0 %)	Variable from low to high level	Can fine-tune expression levels in a dose-dependent manner Tight regulation possible Low basal level Inexpensive inducer
$rhaP_{BAD}$ promoter	L-rhamnose 0.2 %	Variable from low to high level	Tight regulation Low basal activity Relative expensive inducer

Induction can also be achieved by promoters that are activated by starvation of cells for an essential nutrient. For example, phosphate deficiency induces the PhoA promoter in *E. coli* (Sharfstein *et al.*, 1996). Such inducer-free systems allow for inexpensive gene expression, which is particularly important in large-scale fermentations for the production of chemicals and fuels where cost is an issue (Keasling, 2012). On the other hand, the use of such stress-induced promoters do not allow for full control over the moment and intensity of expression (Keasling, 2008).

Bacteriophages have evolved very efficient promoters in order rapidly redirect the host transcriptional machinery towards their genome. Thus, these promoters achieve the highest expression levels in bacteria (Sørensen & Mortensen, 2005; Terpe, 2006). The described regulatory components can be used to control such bacteriophage promoters, which is most commonly achieved by the lactose or IPTG[1]-inducible LacI-system (Sørensen & Mortensen, 2005). The most common promoters used for high-level gene expression are the pL promoter from the lambda-phage (Bernard *et al.*, 1979), the pN25 promoter from phage T5 (Gentz & Bujard, 1985) or the promoter driving expression of the viral DNA-polymerase in phage T7 (Tabor & Richardson, 1985). While the T5 and lambda phages rely entirely on the host transcriptional machinery, the T7 expression system uses the phage's RNA-polymerase that specifically binds to its own promoter.

Inducible gene expression in *S. cerevisiae*

The most commonly used system for tightly regulated gene expression in *S. cerevisiae* is based on the GAL1 promoter (Blazeck *et al.*, 2012; Maya *et al.*, 2008). This promoter is a prominent example for illustrating the complex transcriptional regulation in eukaryotes. In its native environment, the GAL1 promoter drives expression of a galactokinase as part of the yeast galactose utilization pathway. Via a complex metabolic and genetic network consisting of several structural as well as multi-level regulatory components, this promoter is strongly activated by galactose, whereas glucose causes its repression (Giniger *et al.*, 1985; Lamphier & Ptashne, 1992). Switching of these sugars results in a strong induction. However, apart from the high cost of galactose, stringent media requirements restrict the range of possible applications. For example, glucose released from hydrolyzed plant polysaccharides would repress the GAL1 promoter, preventing its use for the microbial conversion of lignocellulose to chemicals.

One way to enable a more independent control over gene expression in *S. cerevisiae* and eukaryotes in general is the use of heterologous regulation mechanisms. Heterologous transcription factors and their DNA binding sites are typically not recognized by the host, which makes them suitable for orthogonal gene regulation that avoids cross-talk with the host metabolism. For example, the native function of the prokaryotic TetR-repressor can be inverted when it is fused to the

[1] Isopropyl β-D-1-thiogalactopyranoside (IPTG) is a lactose derivative and normally used to induce LacI-mediated promoters.

activation domain of a viral transcription factor that is functional in eukaryotes (Garí *et al.*, 1997).

Alternative inducible promoters that are less influenced by external or metabolic conditions have been developed, but these systems are often either weaker, have lower dynamic ranges or require inducer molecules, the costs of which would prohibit their use in large-scale applications (Keasling, 2012; Cartwright et al., 1994; Maya et al., 2008).

For most applications in bacteria and eukaryotes, the properties of an ideal inducible promoter can be summarized as follows: It should be inducible over a wide dynamic range (on/off ratio), with minimal activity in the repressed state and high transcription rates when induced. Such a system should be insensitive to external and metabolic conditions, uniquely guiding expression of the target gene(s), without interfering with the host metabolism. In addition, inducer molecules should be inexpensive, which is especially important in large-scale fermentations of low-value products, such as biofuels or commodity materials (Keasling, 2012; S. K. Lee *et al.*, 2008).

Objectives of this thesis

Find and apply the genetic basis of bacterial ionic liquid tolerance

Knowledge of an IL resistance mechanism would be of scientific interest and potentially useful for the conversion of biomass to fuels and other chemicals. By using a targeted approach that involved functional screening of the *E. lignolyticus* DNA, I pursued my first goal of finding the genetic basis of IL tolerance.

Provide an inducible gene expression system

With the insights gained during the course of my research, I set the objective to work towards an ideal expression system. This system will enable controllable gene expression that is based on the EilR repressor and its DNA binding site.

Chapter 1

IDENTIFICATION AND APPLICATION OF BACTERIAL IONIC LIQUID TOLERANCE

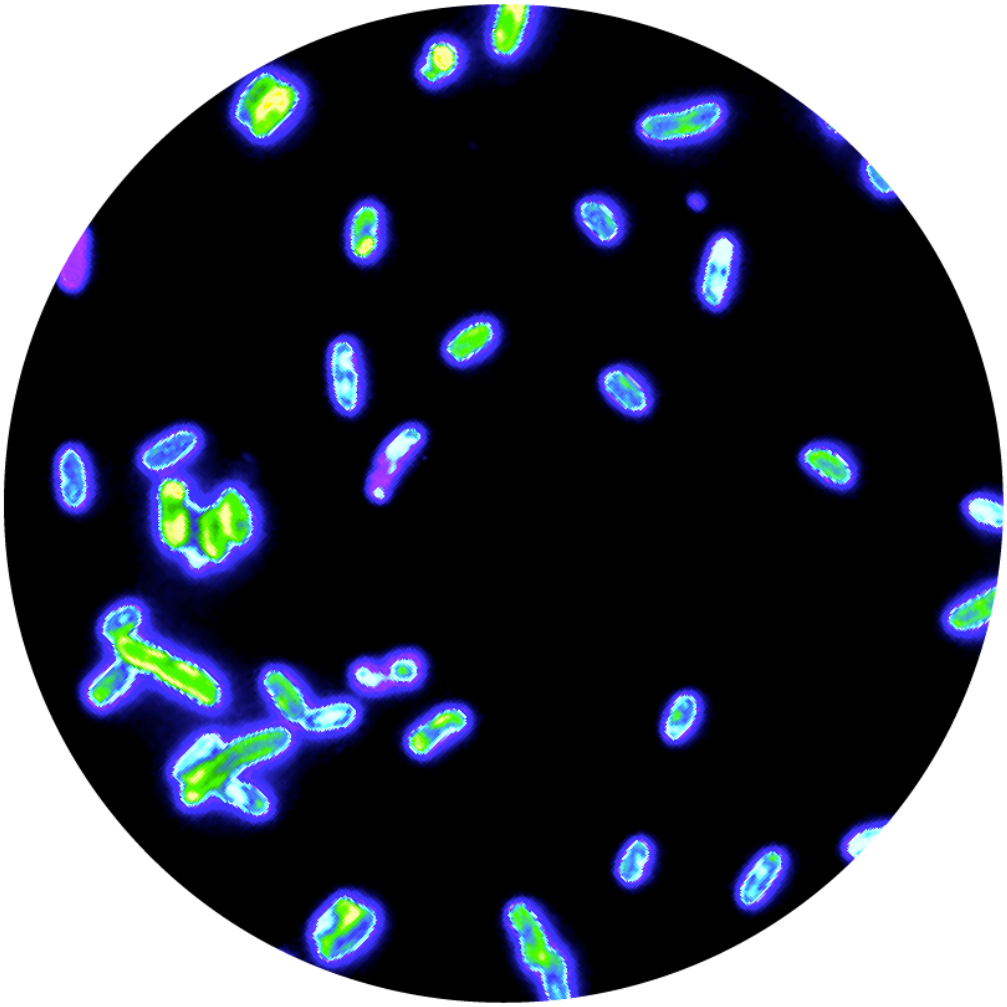


Image: *Escherichia coli* cells expressing the membrane-bound EilA-efflux pump fused to GFP

nature COMMUNICATIONS

ARTICLE

Received 4 Oct 2013 | Accepted 21 Feb 2014 | Published 26 Mar 2014

DOI: 10.1038/ncomms4490 OPEN

An auto-inducible mechanism for ionic liquid resistance in microbial biofuel production

Thomas L. Ruegg[1,2,3], Eun-Mi Kim[1,4], Blake A. Simmons[1,5], Jay D. Keasling[1,4,6], Steven W. Singer[1,4], Taek Soon Lee[1,4] & Michael P. Thelen[1,3]

Ionic liquids (ILs) are emerging as superior solvents for numerous industrial applications, including the pretreatment of biomass for the microbial production of biofuels. However, some of the most effective ILs used to solubilize cellulose inhibit microbial growth, decreasing efficiency in the overall process. Here we identify an IL-resistance mechanism consisting of two adjacent genes from *Enterobacter lignolyticus*, a rain forest soil bacterium that is tolerant to an imidazolium-based IL. These genes retain their full functionality when transferred to an *Escherichia coli* biofuel host, with IL resistance established by an inner membrane transporter, regulated by an IL-inducible repressor. Expression of the transporter is dynamically adjusted in direct response to IL, enabling growth and biofuel production at levels of IL that are toxic to native strains. This natural auto-regulatory system provides the basis for engineering IL-tolerant microbes, which should accelerate progress towards effective conversion of lignocellulosic biomass to fuels and renewable chemicals.

[1] Joint BioEnergy Institute, 5885 Hollis Street, Emeryville, California 94608, USA. [2] Institute of Botany, University of Basel, Basel 4056, Switzerland. [3] Biology and Biotechnology Division, Lawrence Livermore National Laboratory, Livermore, California 94550, USA. [4] Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. [5] Biological and Materials Science Center, Sandia National Laboratories, Livermore, California 94550, USA. [6] Department of Chemical and Biomolecular Engineering, Department of Bioengineering, University of California, Berkeley, California 94720, USA. Correspondence and requests for materials should be addressed to M.P.T. (email: mthelen@llnl.gov).

Lignocellulosic biomass is an abundant resource that is available for the sustainable production of biofuels and high-value chemicals. Some of the most promising approaches have centered on engineering microbes[1–3] and utilizing a wide range of feedstocks, including woody biomass, indigenous grasses and agricultural residues such as corn stover[4,5]. The inherent recalcitrance of biomass requires an initial pretreatment step to render polysaccharides free from lignin for subsequent enzymatic or chemical hydrolysis to fermentable sugars. To solubilize lignocellulosic biomass, certain hydrophilic ILs are highly effective and environmentally friendly pretreatment agents that generate relatively low amounts of biomass-derived inhibitors compared with other conventional pretreatment methods[6–9]. A major disadvantage of the commonly used imidazolium ILs is their intrinsic microbial toxicity, which impairs growth of typical biofuel-producing hosts such as *Escherichia coli* and *Saccharomyces cerevisiae,* hence preventing efficient biofuel production[10–12]. In addition, the inhibition of biofuel synthetic enzymes by these ILs can severely reduce the yield of the final product[11].

It was recently demonstrated that an engineered *E. coli* strain is able to convert IL-pretreated biomass into biofuels in laboratory-scale experiments[5]. However, the extensive washing required for complete IL removal is not feasible in large-scale, industrial applications. An ideal and more sustainable process should balance the costs of removing IL with fermentation performance[13]. A novel way to achieve this would use biofuel-producing microbes that can tolerate residual levels (for example, 0.2–5% wt/vol) of ILs.

Because a mechanism for microbial tolerance to ILs has not been elucidated, we investigated an IL-tolerant bacterium isolated from rain forest soils. *Enterobacter lignolyticus* grows well in at least 380 mM (5.5%) 1-ethyl-3-methylimidazolium chloride (abbreviated as [C_2mim]Cl)[14,15], an IL that is a promising pretreatment solvent for biomass[6]. Using RNA-sequencing transcriptomics, we determined that *E. lignolyticus* responds to [C_2mim]Cl exposure by the differential expression of 688 genes, resulting in a complex response with numerous phenotypic changes[15]. These include an increased production of cyclopropane fatty acids, scavenging of compatible solutes and regulation of membrane transporters.

Here we use a targeted functional screening approach, as a complement to the global nature of transcriptomics, to identify the key genetic elements responsible for [C_2mim]Cl tolerance and discover that these consist solely of an efflux pump and its regulator. We transfer these genes to an engineered *E. coli* host, and demonstrate enhanced production of a terpene-based biofuel in the presence of this IL.

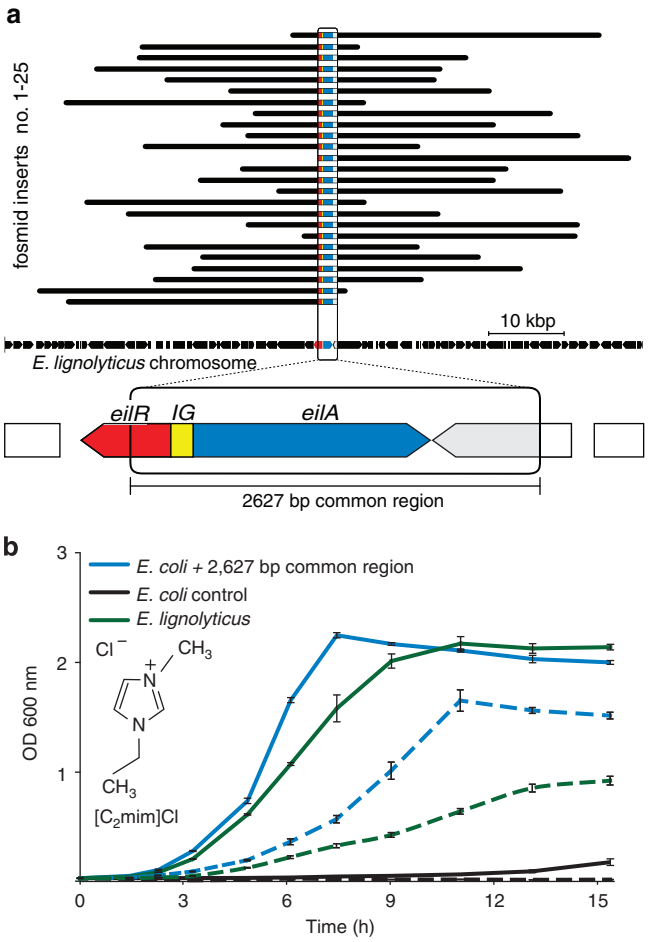


Figure 1 | Identification of *eilA* by its function in *E. coli*. (**a**) An *E. coli* fosmid library containing fragments of genomic DNA from *E. lignolyticus* was screened for growth on [C_2mim]Cl by plating cells directly following phage infection. After end-sequencing from the fosmid insertion sites of 25 [C_2mim]Cl-tolerant clones, these sequences (black bars) were aligned with the genome of *E. lignolyticus*. All inserts shared 2,627-bp (boxed region) containing the complete *eilA* (blue bar) and an adjacent section of the *eilR* regulatory gene (red bar). The *eilAR* module comprises these two genes and the 117-bp intergenic region (*IG*; yellow bar). (**b**) Growth of *E. lignolyticus* (green curves), and *E. coli* containing a plasmid, either without (black) or with the common region (blue) in 272 mM (solid line) and 410 mM (dashed line) [C_2mim]Cl. The chemical structure of [C_2mim]Cl is shown within the graph. The curves and error bars represent the means and standard deviation of biological triplicate measurements.

Results

A single gene is responsible for IL tolerance. To cover a potentially complex IL-tolerance mechanism with many genes involved, we used relatively large DNA fragments (30–50 kb) to construct a fosmid library for screening in *E. coli*. Out of ~6,000 clones, 41 (0.7%) evinced an IL-tolerance phenotype by growing rapidly in the presence of 136 mM [C_2mim]Cl. Sequence analysis of these clones revealed a common region of 2.6 kb encompassing three genes, only one of which was entirely represented (Fig. 1a). *E. coli* carrying this common region grew at rates similar to or greater than those of the source bacterium *E. lignolyticus* in a medium containing up to 410 mM [C_2mim]Cl (Fig. 1b; Supplementary Fig. 1). This protective effect was not due to catabolism of the IL, as there was no measurable decrease in the [C_2mim]$^+$ cation concentration after culturing cells to the stationary phase compared with the beginning of culture (Supplementary Fig. 2).

The complete gene, designated *eilA* (Enterobacter IL tolerant; locus tag Entcl_2352), encodes a 52-kDa protein that is homologous to the ubiquitous major facilitator superfamily (MFS) of membrane transporters, although no orthologue is present in *E. coli*. The identification of this gene corroborates our previous observation that it is one of the most highly upregulated transcripts in *E. lignolyticus* when exposed to [C_2mim]Cl[15]. Sequence analysis indicated that EilA is a homologue of proton antiporters, possessing 14 transmembrane helices that span the inner membrane. Numerous representatives of this MFS subgroup are responsible for prokaryotic and eukaryotic multidrug resistance, capable of transporting a range of hydrophobic cations, including some widely used bactericidal quaternary ammonium compounds and dyes[16–18]. Assuming a similar function of EilA, we screened a library consisting of 240 compounds representing many different structures and properties

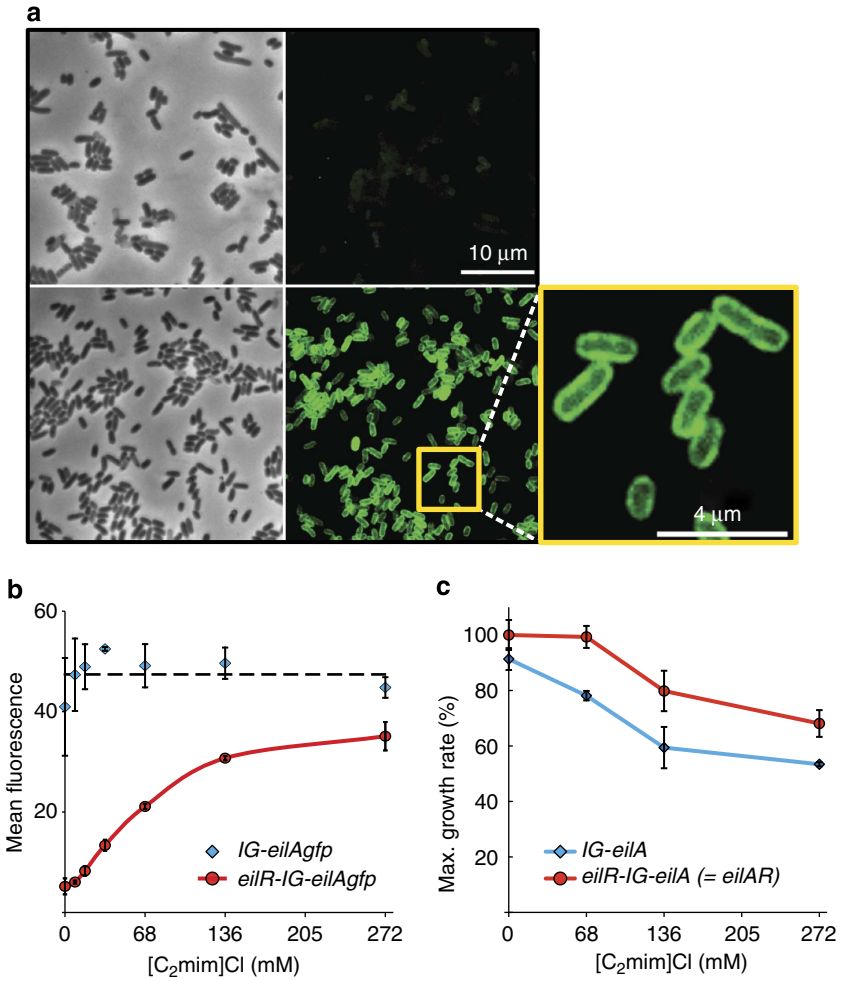


Figure 2 | [C_2mim]Cl-inducible repressor EilR regulates *eilA* expression. (**a**) Phase contrast (left) paired with fluorescence (right) microscopy images of *E. coli* cells expressing an EilA–GFP fusion protein from a plasmid carrying the *eilR–IG–eilAgfp* cassette. Cultures were grown in the absence of [C_2mim]Cl (upper) or in the presence of 272 mM [C_2mim]Cl (lower). The magnification shows membrane localization of EilA–GFP. (**b**) Mean single-cell fluorescence of *E. coli* expressing EilA–GFP from the native promoter, either constitutively (*IG–eilAgfp*: blue diamonds) or under the control of EilR (*eilR–IG–eilAgfp*: red circles). Cells were grown in the indicated [C_2mim]Cl concentrations before flow cytometry measurements. The dashed line indicates the general trend. (**c**) Maximum growth rates of *E. coli* expressing *eilA* via the native promoter, either constitutively (blue diamonds) or under the control of EilR (red circles). For both **b** and **c**, the curves and error bars represent the means and standard deviation of biological triplicate measurements.

and found that EilA exports primarily hydrophobic ammonium cations (Supplementary Data 1; Supplementary Fig. 3). It is noteworthy that *E. coli* expressing *eilA* is relatively resistant to methyl viologen (paraquat), a divalent bipyridinium cation. This observation is consistent with a group of efflux pumps that share up to 77% sequence identity to EilA, and which export methyl viologen from other Enterobacteriaceae, such as *Salmonella* and *Klebsiella* species[19,20].

We also examined the tolerance of an *E. coli eilA* strain to other imidazolium-based ILs, comparing different alkyl chain lengths at the N1 position on $[C_2mim]^+$ and also replacing the Cl^- anion with acetate. In general terms, we observed a correlation between an increase in alkyl chain lengths and toxicity, and that acetate was tolerated but to a lesser extent than Cl^- (Supplementary Figs 1 and 4).

Expression is regulated by a [C_2mim]Cl-inducible repressor. In nature, bacteria are exposed to fluctuating concentrations of a broad spectrum of organic compounds. To respond to potentially toxic effects, expression of the appropriate transporters is typically adjusted by substrate-inducible transcriptional regulators[21]. The *E. lignolyticus eilR* gene (locus tag Entcl_2353), located upstream of *eilA* and only partially present on the 2.6-kb region, encodes a 22-kDa protein of the TetR family of transcriptional repressors (Fig. 1a). Proteins of this family are known to be prominent regulators of bacterial multidrug efflux pumps, including MFS transporters[22]. The 117-bp noncoding intergenic region (*IG*; Fig. 1a) contains the predicted promoter and operator sequences for the two divergently transcribed genes. In experiments with *E. coli* that included only the *IG-eilA* sequence, [C_2mim]Cl tolerance was maintained, confirming that the *eilA* promoter sequence resides within *IG* and is operational in *E. coli* (Supplementary Fig. 5). To analyse the regulation of *eilA* expression, we measured the fluorescence of an EilA–green fluorescent protein (GFP) fusion protein, which confirmed that expression from the native promoter is constitutive. Introducing *eilR* into the cassette resulted in repression in the absence of [C_2mim]Cl, whereas increasing the [C_2mim]Cl level directly correlated with EilA–GFP expression (Fig. 2a,b). Therefore, we postulate that [C_2mim]Cl induces the release of EilR from its operator site, resulting in activation of *eilA* transcription. A similar mechanism has been demonstrated by several studies on homologous Tet repressor-regulated efflux pump systems[22], including the methyl viologen operon *pqrAB* in *Streptomyces*[23].

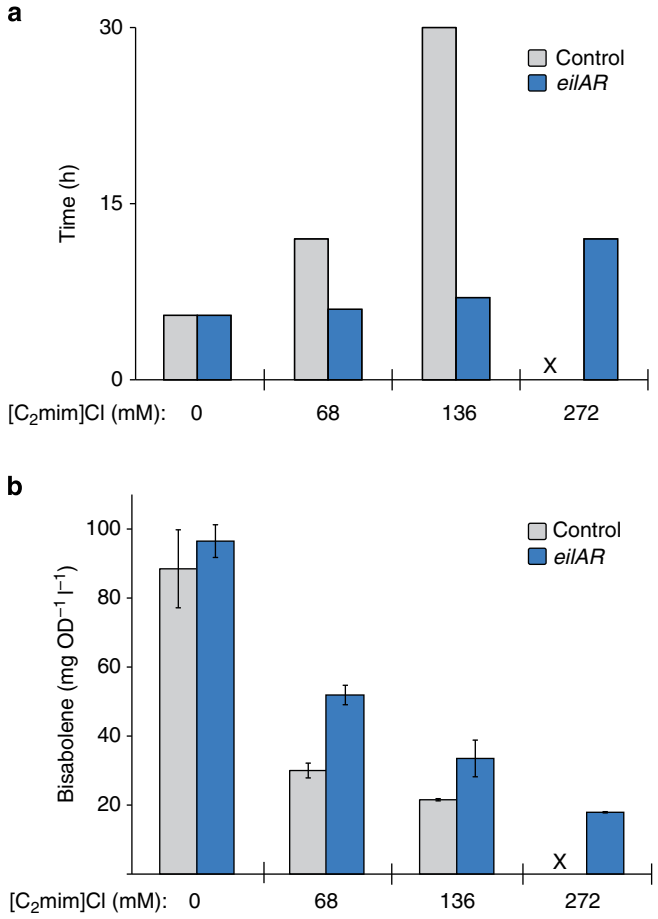


Figure 3 | Bacterial bisabolene production in media containing [C$_2$mim]Cl. *E. coli* harbouring a plasmid containing the bisabolene pathway and a second plasmid either with the *eilAR* cassette (blue bars) or lacking an insert as control (grey bars). (**a**) Time required for each culture to reach $OD_{600nm}=1$, the standard cell density used for inducing the bisabolene pathway. (**b**) Bisabolene production measured 6 days after induction of cultures grown in three [C$_2$mim]Cl concentrations, as indicated with columns and error bars representing the means and standard deviation of technical triplicate measurements. Values were normalized to cell density to illustrate productivity of the bacteria. X indicates that no growth or production was observed.

In the presence of 0.1 mM methyl viologen, *E. coli* expressing *eilA* constitutively grew at a near-normal rate. However, insertion of *eilR* in its native alignment (as *eilAR*) resulted in complete growth inhibition. Moreover, when *E. lignolyticus* was exposed to methyl viologen, the lag phase was considerably prolonged (Supplementary Fig. 6), indicating that EilR is not induced by this EilA substrate. Thus, in other phenotypic screening approaches, candidate genes such as those encoding efflux pumps could be overlooked when expression is restricted by substrate-specific transcriptional regulation.

Overexpression of a transporter can overload the cell membrane and have deleterious consequences on cell viability[24], which we observed when *eilA* was expressed at higher levels from an isopropyl-β-D-thiogalactoside-inducible promoter (Supplementary Fig. 7). To prevent this while maintaining maximum export capacity of the inhibitory compound, a substrate-responsive efflux system has been suggested for biotechnological applications such as those used in biofuel production[25]. We confirmed this by testing the growth of *E. coli* strains over a range of [C$_2$mim]Cl levels. Compared with a strain in which *eilA* is expressed constitutively under the control of its native promoter, higher growth rates were observed in the presence of the repressor (Fig. 2c). We reasoned that EilR increases cell viability by adjusting efflux pump expression to [C$_2$mim]Cl concentration, so we included the entire native *eilAR* cassette in experiments using a biofuel-producing *E. coli* strain.

Engineering an IL-tolerant biofuel production strain. Bisabolane is an advanced biofuel candidate with combustion properties comparable to diesel. The dehydrogenated precursor, bisabolene, is produced from acetyl-coenzyme A (CoA) in *E. coli* with a two-plasmid system: one plasmid contains eight heterologous mevalonate pathway genes, and the other has the plant-derived bisabolene synthase gene[26]. We assembled these genes onto a single plasmid optimized for bisabolene production, then introduced a second plasmid containing the *eilAR* cassette to create a prototype IL-tolerant biofuel production system. To represent different possible fermentation scenarios, we tested cell growth and bisabolene production by this engineered *E. coli* strain in a medium containing a range of [C$_2$mim]Cl concentrations from 0 to 270 mM, representing the residual [C$_2$mim]Cl remaining with the hydrolysate after pretreatment and saccharification[13].

For *E. coli* lacking *eilAR*, increasing the [C$_2$mim]Cl concentration in the growth medium caused a marked increase in lag time, up to 445% at 136 mM, and no growth was observed at 272 mM (Fig. 3a; Supplementary Fig. 1). Even at 63 mM [C$_2$mim]Cl, the lowest level tested, a doubling in lag time would hamper the overall efficiency of a biofuel production process[3]. In contrast to the control, we found the growth of cultures expressing *eilAR* only slowed moderately with increasing [C$_2$mim]Cl, to a maximum lag time increase of 120% at 272 mM.

Cultures were grown to the same cell density before inducing the biosynthetic pathway. Bisabolene production of both the *eilAR* and the control strain was similar in [C$_2$mim]Cl-free medium, whereas the *eilAR* cassette enabled increased productivity over the control in the presence of [C$_2$mim]Cl. In 68 mM [C$_2$mim]Cl, cells expressing *eilAR* produced 52 mg $OD_{600}^{-1}l^{-1}$ bisabolene, 73% more than the control strain. This effect was apparent with increasing IL concentrations (Fig. 3b, Supplementary Fig. 8). We observed decreased bisabolene production for both strains in [C$_2$mim]Cl compared with the IL-free medium. Productivity decreased by 66% in the control and 46% in the *eilAR* strain in 68 mM [C$_2$mim]Cl. Certain ILs have been shown to interfere with production of biofuels at levels below the growth inhibition threshold, most likely owing to direct inhibition of biosynthetic pathway enzymes. For example, bacterial biodiesel yield decreased by 25% in the presence of only 6 mM of an imidazolium-based IL[11].

Discussion

In summary, we have shown that targeted, functional screening of DNA from environmental microorganisms is a powerful approach to discover the mechanisms underpinning phenotypes such as bacterial resistance to inhibitory compounds. Such mechanisms can include regulatory proteins that specifically respond to natural and synthetic inducer molecules. The sequence-independent method used here can be expanded to detect further mechanisms useful for engineering other IL-sensitive host organisms such as *Clostridium spp.* or *Saccharomyces cerevisiae*.

In our case, transferring a genetic module from a rain forest soil microbe to an *E. coli* biofuel host conferred the tolerance needed for it to grow well in the presence of toxic concentrations of ILs. We demonstrated that this introduced efflux mechanism

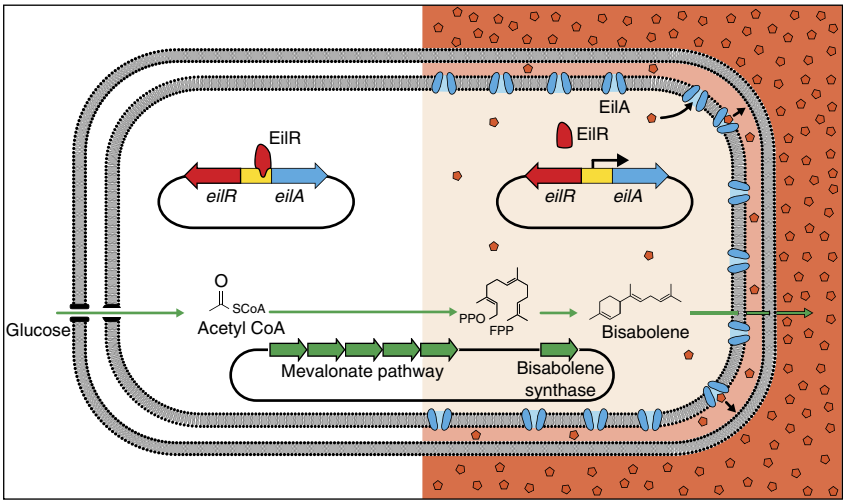


Figure 4 | Model of the ionic liquid-tolerant bacterium converting hydrolysed biomass into biofuel. The cell harbours a plasmid containing the bisabolene biosynthetic genes (green), and another plasmid with the *eilAR* cassette encoding the efflux pump EilA (blue) and repressor EilR (red). Sugars from the hydrolysate are metabolized to acetyl-CoA, which then feeds into the biofuel production pathway (green arrows). In the absence of ionic liquid (left half), EilR binds to the operator in the IG (yellow), therefore preventing *eilA* transcription. The presence of the cellulose solubilizing ionic liquid [C_2mim]Cl (orange pentagons; right half) causes EilR to release from the operator site, resulting in EilA expression and membrane insertion. EilA then exports toxic [C_2mim]Cl from the cell, enabling robust growth and efficient production of biofuels.

improved production of a biofuel precursor in the presence of low (residual) levels of an IL used in biorefineries (Fig. 4). We anticipate that engineering IL-tolerant biofuel pathway enzymes and production strains with tolerance to inhibitors originating from biomass breakdown[27] are needed to further increase yields. Autoregulation of IL-efflux by the EilR repressor increased strain robustness and is particularly important for efficient production in industrial-scale fermentations, where residual IL levels fluctuate between biomass batches. In addition, such a substrate-responsive system obviates the need for costly inducer molecules. An IL-tolerant biofuel host could also convert the adverse effects of these solvents into an advantage by preventing growth of microbial contaminants, thus allowing fermentation under more economical, aseptic conditions. By eliminating a bottleneck of a promising biomass pretreatment method, our findings could contribute to an effective and sustainable production of biofuels and chemicals.

Methods

Reagents. All chemicals were purchased from Sigma-Aldrich.

Fosmid library construction and screening. To obtain genomic DNA, *E. lignolyticus* was cultivated in 10 ml Luria Bertani medium for 16 h at 30 °C, then centrifuged and cells resuspended in TE buffer to an OD_{600} of 1. Lysozyme (1.7 mg ml^{-1} final concentration) was added to the cell suspension and incubated for 30 min at 37 °C. Following this, a standard cetyltrimethylammonium bromide method was used to extract genomic DNA[28]. The genomic DNA library was constructed in the pCC1FOS fosmid vector (Epicentre biotechnologies), using the manufacturer's protocol, but the size separation step was omitted as the extracted DNA had the average size needed for fosmid cloning (~40 kb). A control library was built using 42-kb human control DNA. The titre of phage particles was determined for both *E. lignolyticus* and control libraries and diluted to receive ~1,500 clones per 120 mm × 120 mm plate. Infected *E. coli* EPI300 cells were spread directly on Luria Bertani agar plates containing 12.5 mg l^{-1} chloramphenicol and 136 mM [C_2mim]Cl and incubated at 37 °C. After 20 h, the clearly detectable tolerant clones were selected from all four plates and grown overnight in media containing 0.01% arabinose to induce a high fosmid copy number. Fosmids were isolated using standard column DNA purification (Qiagen). The insert ends were sequenced using the suggested primers pCC1FOS-F (5′-GGATGTGCTGCAAGGCGATTAAGTTGG-3′) and pCC1FOS-R (5′-CTCGTATGTTGTGTGGAATTGTGAGC-3′). The sequences (600–1,000 bp) were then aligned to the *E. lignolyticus* genome (GenBank ID: CP002272) to determine the insert sequences between the ends.

Plasmid and strain construction. *E. coli* DH10B and DH1 were used for bacterial transformation and plasmid storage. A [C_2mim]Cl-tolerant fosmid clone was used as a template to amplify *E. lignolyticus* DNA. Most plasmids were constructed from BglBrick vectors[29] using restriction enzymes. For plasmids containing the truncated intergenic region (pTR-9–pTR-12), the PCR product, which included the vector backbone, was phosphorylated before blunt-end self-ligation. Using circular polymerase extension cloning[30], a 30-bp linker sequence (5′-GCTGGCTCTGCTGCAGGTTCTGGTGAATTT-3′) together with the superfolder *gfp*[31] was fused to the *eilA* 3′-end before the stop codon. *eilA* was expressed from low-copy plasmids (SC101), with the exception of the growth rate experiment (Fig. 2c), where a high-copy plasmid (ColE1) was used. An overview of primers and plasmids is given in Supplementary Tables 1 and 2. For the production of bisabolene, plasmids were prepared according to the BglBrick cloning strategy[29]. The top portion (MevT) of the mevalonate pathway contains genes for the conversion of acetyl-CoA to mevalonate: acetoacetyl-CoA synthase from *E. coli* (AtoB), HMG-CoA synthase from *S. cerevisiae*, and an amino-terminal truncated HMG-CoA reductase from *S. cerevisiae*. The bottom portion (MBIS) of the mevalonate pathway contains genes for the conversion of mevalonate to FPP: mevalonate kinase from *S. cerevisiae*, phosphomevalonate kinase from *S. cerevisiae*, phosphomevalonate decarboxylase from *S. cerevisiae*, IPP isomerase from *E. coli* (idi) and farnesyl diphosphate synthase from *E. coli* (ispA). The codon optimized (co) (E)-γ-bisabolene synthase from *Abies grandis* was used[26]. Production vector pBbA5c-MevT(co)-MBIS(co)-T_{1002}-P_{trc}-Bis(co) (JBEI-6523) was constructed using JBEI-3100 (pBbA5c-MevT(co)) as a vector (*Bam*HI/*Xho*I) and JBEI-4174 (pBbS5k-MBIS-(co)-T_{1002}-P_{trc}-Bis(co)) as an insert (*Bgl*II/*Xho*I).

Growth experiments. *E. lignolyticus* or freshly transformed *E. coli* MG1655 (except where indicated) were grown to the stationary phase. These seed cultures were then diluted 1:200 into EZ-Rich medium (Teknova), supplemented with 1% glucose and the appropriate amounts of antibiotics (ampicillin 100 mg l^{-1}, chloramphenicol 12.5 mg l^{-1} and kanamycin 50 mg l^{-1}). Cultures were grown at 37 °C in 24-well microtitre plates on the Infinite F-200 (Tecan) or Synergy 4 (BioTek) readers, or where indicated, in 200-ml baffled flasks at 200 r.p.m. Optical density was measured at a wavelength of 600 nm. The calculation of growth rates is based on the curve-fitting model of Baranyi and Roberts, which generates nearly linear growth rates in the mid-log phase[32].

Omnilog inhibitor screen. Seed cultures of *E. coli* MG1655 containing either pTR_2 or pBbS0c as control, were diluted 1:200 into a EZ-rich medium supplemented with 12.5 mg l^{-1} chloramphenicol and a 1:100 dilution of redox dye A (Biolog). This stock culture was used to inoculate plates PM11 to PM20 at 100 µl per well. The plates were incubated for 22 h at 37 °C in an Omnilog plate reader. Colour changes, resulting from reduction in the redox dye, were recorded in each well. These kinetic data were analysed with OmniLog PM software (Biolog).

Flow cytometry and fluorescence microscopy. Stationary phase cultures expressing the EilA–GFP fusion protein were diluted 1:1,000 in PBS buffer.

Using a Guava easyCyte (Millipore) flow cytometer, cells were counted (2,000 events per sample) by forward and side-scatter acquisition, and the cellular accumulation of EilA–GFP was measured by fluorescence intensity. Data acquisition was performed using InCyte software version 2.2 (Millipore). A strain of *E. coli* lacking GFP was used to subtract background fluorescence. For microscopy, cells were immobilized on a slide covered with a thin film of 1.5% agarose. Microscopy was performed on a Leica GM4000B fluorescence microscope. Images were taken with an
Orca-HR camera (Hamamatsu) and processed with the MetaMorph 7.7 software (Molecular Devices).

Bisabolene production. *E. coli* DH1 was co-transformed with pJBEI-6523 (bisabolene pathway) and either pTR_8 (*eilAR)* or pBbS0a (control) and grown to the stationary phase. As a 1:100 dilution, these strains were grown at 37 °C (200 r.p.m.) in baffled flasks containing 50 ml EZ-rich medium (1% glucose, ampicillin 100 mg l^{-1} and chloramphenicol 30 mg l^{-1}) and the appropriate amounts of [C_2mim]Cl to reach an optical density (OD_{600nm}) of 1 for induction with 100 μM isopropylthiogalactoside (3IPTG). Three 8 ml aliquots were transferred to culture tubes and overlaid with 10% dodecane as the organic phase for bisabolene extraction. Ten microlitres of the dodecane overlays were sampled and diluted into 990 μl of ethyl acetate spiked with caryophyllene as an internal standard. The samples were analysed by GC/MS (Agilent 6,890 gas chromatography/5,973 mass spectrometry). The oven temperature was initiated at 100 °C for 0.75 min with a temperature gradient up to 250 °C at a rate of 40 °C min^{-1}, and held for 1 min at 250 °C. Injector and MS quadrupole detector temperatures were 230 °C and 150 °C, respectively. The MS was operated in a selected ion monitoring mode using fragment ions m/z 161, 189 and 204 for bisabolene identification and quantification.

Quantification of [C_2mim$^+$] by mass spectrometry. *E. coli* harbouring an *eilAR* plasmid was grown to the stationary phase in M9 minimal medium containing 0.4% glucose and [C_2mim]Cl at 0 mM, 68 mM, 136 mM or 272 mM. Media was collected before and after cultivation, centrifuged and equally diluted in 20 mM ammonium acetate and 0.2% formic acid in methanol:water (20:80 by volume) to a final [C_2mim]Cl concentration between 5 and 50 μM. Samples (50 μl) were injected at a flow rate of 0.3 ml min^{-1}. Detection of $[C_2mim]^+$ (m/z = 111) was carried out using an Agilent 1,100 HPLC (Agilent Technologies Inc., Santa Clara, CA, USA) and API 2000 LC/MS/MS system (AB SCIEX, Foster city, CA, USA). Measurements were performed using a previously described method[33] with the following modifications: We used rapid direct infusion electrospray ionization mass spectrometry in the positive scan and selected ion mode. Also, our electrospray ionization conditions were 4.5-kV spray voltage with an auxiliary gas (N_2) flow of 20 and 8 (arbitrary units), and a heated ion transfer capillary/mass spectrometer inlet temperature of 350 °C.

References

1. Rabinovitch-Deere, C. A., Oliver, J. W. K., Rodriguez, G. M. & Atsumi, S. Synthetic biology and metabolic engineering approaches to produce biofuels. *Chem. Rev.* **113,** 4611–4632 (2013).
2. Wen, M., Bond-Watts, B. B. & Chang, M. C. Production of advanced biofuels in engineered E. coli. *Curr. Opin. Chem. Biol.* **17,** 472–479 (2013).
3. Liu, T. G. & Khosla, C. Genetic engineering of Escherichia coli for biofuel production. *Annu. Rev. Genet.* **44,** 53–69 (2010).
4. de Jong, B., Siewers, V. & Nielsen, J. Systems biology of yeast: enabling technology for development of cell factories for production of advanced biofuels. *Curr. Opin. Biotechnol.* **23,** 624–630 (2012).
5. Bokinsky, G. *et al.* Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered Escherichia coli. *Proc. Natl Acad. Sci. USA* **108,** 19949–19954 (2011).
6. Mora-Pale, M., Meli, L., Doherty, T. V., Linhardt, R. J. & Dordick, J. S. Room temperature ionic liquids as emerging solvents for the pretreatment of lignocellulosic biomass. *Biotechnol. Bioeng.* **108,** 1229–1245 (2011).
7. Liu, C. Z., Wang, F., Stiles, A. R. & Guo, C. Ionic liquids for biofuel production: opportunities and challenges. *Appl. Energ.* **92,** 406–414 (2012).
8. Blanch, H. W., Simmons, B. A. & Klein-Marcuschamer, D. Biomass deconstruction to sugars. *Biotechnol. J.* **6,** 1086–1102 (2011).
9. Li, C. L. *et al.* Comparison of dilute acid and ionic liquid pretreatment of switchgrass: biomass recalcitrance, delignification and enzymatic saccharification. *Bioresour. Technol.* **101,** 4900–4906 (2010).
10. Ouellet, M. *et al.* Impact of ionic liquid pretreated plant biomass on Saccharomyces cerevisiae growth and biofuel production. *Green Chem.* **13,** 2743–2749 (2011).
11. Park, J. I. *et al.* A thermophilic ionic liquid-tolerant cellulase cocktail for the production of cellulosic biofuels. *PLoS One* **7,** e37010 (2012).
12. Nancharaiah, Y. V. & Francis, A. J. Alkyl-methylimidazolium ionic liquids affect the growth and fermentative metabolism of Clostridium sp. *Bioresour. Technol.* **102,** 6573–6578 (2011).
13. Klein-Marcuschamer, D., Simmons, B. A. & Blanch, H. W. Techno-economic analysis of a lignocellulosic ethanol biorefinery with ionic liquid pre-treatment. *Biofuel Bioprod. Bior.* **5,** 562–569 (2011).
14. Deangelis, K. M. *et al.* Complete genome sequence of "Enterobacter lignolyticus" SCF1. *Stand. Genomic Sci.* **5,** 69–85 (2011).
15. Khudyakov, J. I. *et al.* Global transcriptome response to ionic liquid by a tropical rain forest soil bacterium, Enterobacter lignolyticus. *Proc. Natl Acad. Sci. USA* **109,** E2173–E2182 (2012).
16. Fluman, N. & Bibi, E. Bacterial multidrug transport through the lens of the major facilitator superfamily. *Biochim. Biophys. Acta* **1794,** 738–747 (2009).
17. Paulsen, I. T., Brown, M. H. & Skurray, R. A. Proton-dependent multidrug efflux systems. *Microbiol. Rev.* **60,** 575–608 (1996).
18. Saier, M. H. & Paulsen, I. T. Phylogeny of multidrug transporters. *Semin. Cell Dev. Biol.* **12,** 205–213 (2001).
19. Ogawa, W., Koterasawa, M., Kuroda, T. & Tsuchiya, T. KmrA multidrug efflux pump from Klebsiella pneumoniae. *Biol. Pharm. Bull.* **29,** 550–553 (2006).
20. Santiviago, C. A. *et al.* The Salmonella enterica sv. Typhimurium smvA, yddG and ompD (porin) genes are required for the efficient efflux of methyl viologen. *Mol. Microbiol.* **46,** 687–698 (2002).
21. Wade, H. MD recognition by MDR gene regulators. *Curr. Opin. Struc. Biol.* **20,** 489–496 (2010).
22. Ramos, J. L. *et al.* The TetR family of transcriptional repressors. *Microbiol. Mol. Biol. Rev.* **69,** 326–356 (2005).
23. Cho, Y. H. *et al.* The pqrAB operon is responsible for paraquat resistance in Streptomyces coelicolor. *J. Bacteriol.* **185,** 6756–6763 (2003).
24. Wagner, S. *et al.* Consequences of membrane protein overexpression in Escherichia coli. *Mol. Cell. Proteomics* **6,** 1527–1550 (2007).
25. Dunlop, M. J. *et al.* Engineering microbial biofuel tolerance and export using efflux pumps. *Mol. Syst. Biol.* **7,** 487 (2011).
26. Peralta-Yahya, P. P. *et al.* Identification and microbial production of a terpene-based advanced biofuel. *Nat. Commun.* **2,** 483 (2011).
27. Klinke, H. B., Thomsen, A. B. & Ahring, B. K. Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. *Appl. Microbiol. Biotechnol.* **66,** 10–26 (2004).
28. Wilson, K. Preparation of Genomic DNA from Bacteria. *Curr. Protoc. Mol. Biol.* Chapter 2, Unit 2.4 (2001).
29. Lee, T. S. *et al.* BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J. Biol. Eng.* **5,** 12 (2011).
30. Quan, J. & Tian, J. Circular polymerase extension cloning of complex gene libraries and pathways. *PLoS One* **4,** e6441 (2009).
31. Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C. & Waldo, G. S. Engineering and characterization of a superfolder green fluorescent protein. *Nat. Biotechnol.* **24,** 79–88 (2006).
32. Baranyi, J. & Roberts, T. A. A dynamic approach to predicting bacterial growth in food. *Int. J. Food Microbiol.* **23,** 277–294 (1994).
33. Stepnowski, P. *et al.* Reversed-phase liquid chromatographic method for the determination of selected room-temperature ionic liquid cations. *J. Chromatogr. A* **993,** 173–178 (2003).

Acknowledgements

The authors thank Adrienne McKee for intellectual input; Jane Khudyakov for the *E. lignolyticus* isolate; Greg Bokinsky, Richard Heins and Aindrila Mukhopadhyay for scientific advice; Hannah Woo for assistance with the Omnilog measurements; William Holtz for assistance with microscopy; Guillaume Cambray for the *gfp* plasmid; Noppadon Sathitsuksanoh for assistance with MS measurements; and Dominique Loque and Urs Ruegg for comments on the manuscript. T.L.R. is grateful to Professor Thomas Boller for supporting independent graduate research abroad. T.L.R. and M.P.T. have a patent application related to this research. Work performed at the Joint BioEnergy Institute (http://www.jbei.org) was funded by the Office of Science, Office of Biological and Environmental Research, of the US Department of Energy under contract DE-AC02-05CH11231. T.L.R. received additional student support from the Emilia Guggenheim-Schnurr Foundation and Freiwillige Akademische Gesellschaft Basel.

Author contributions

T.L.R. and M.P.T. conceived the study, analysed data and wrote the paper. T.L.R. performed all IL-tolerance experiments. E.-M.K. and T.S.L. designed and performed the bisabolene experiments. B.A.S., T.S.L., E.-M.K., S.W.S. and J.D.K. contributed ideas and editing, and all authors approved the final manuscript.

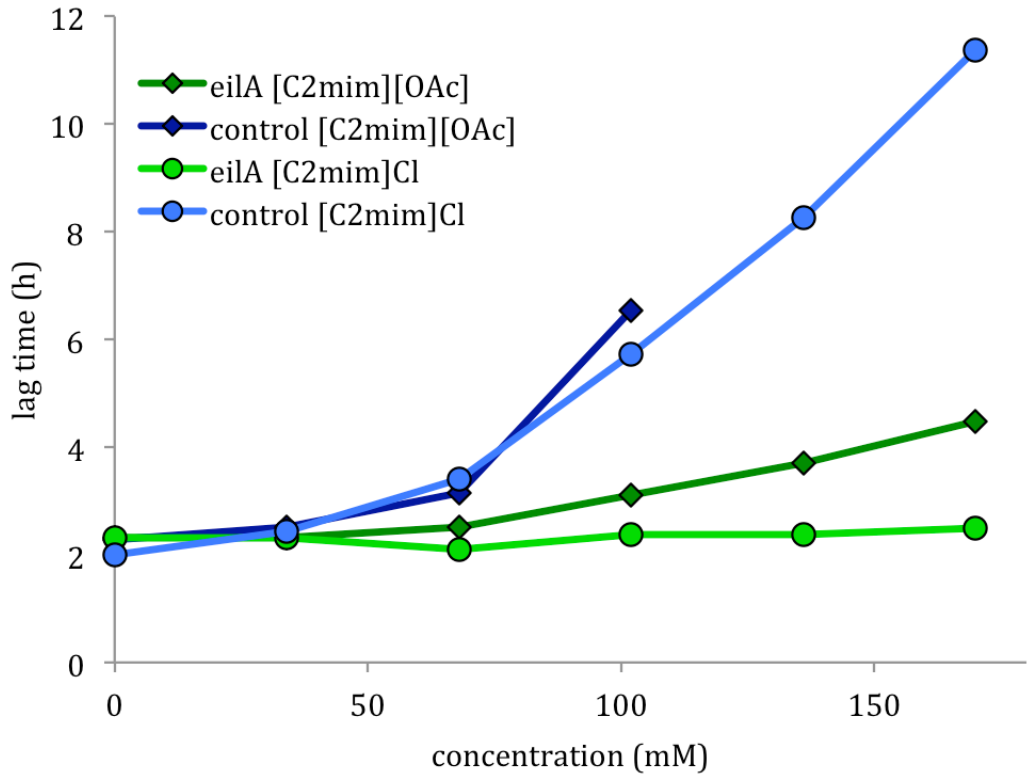
Additional information

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

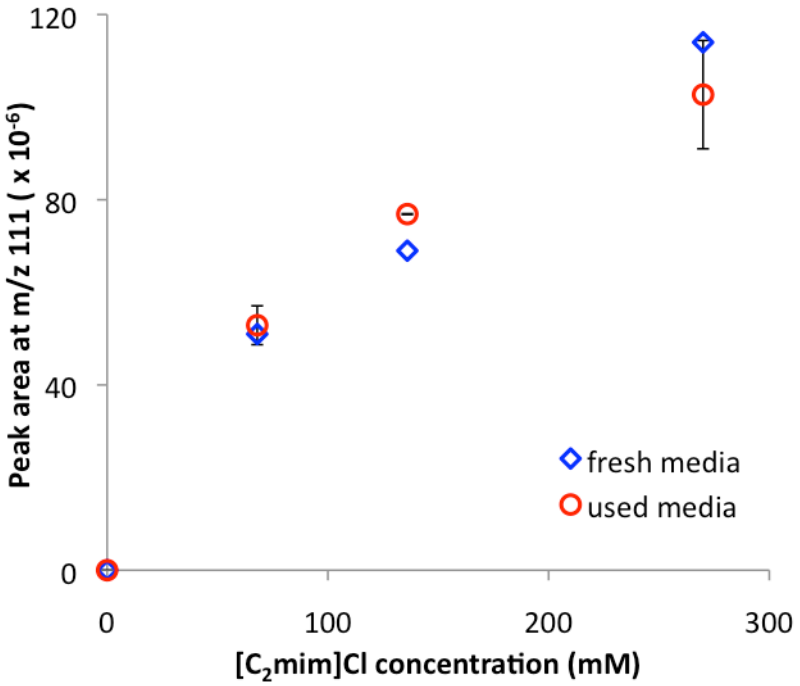
Competing financial interests: J.D.K. has financial interest in Amyris Biotechnologies, LS9 Inc. and Lygos Inc. The remaining authors declare no competing financial interests.

How to cite this article: Ruegg, T. L. *et al.* An auto-inducible mechanism for ionic liquid resistance in microbial biofuel production. *Nat. Commun.* 5:3490 doi: 10.1038/ncomms4490 (2014).

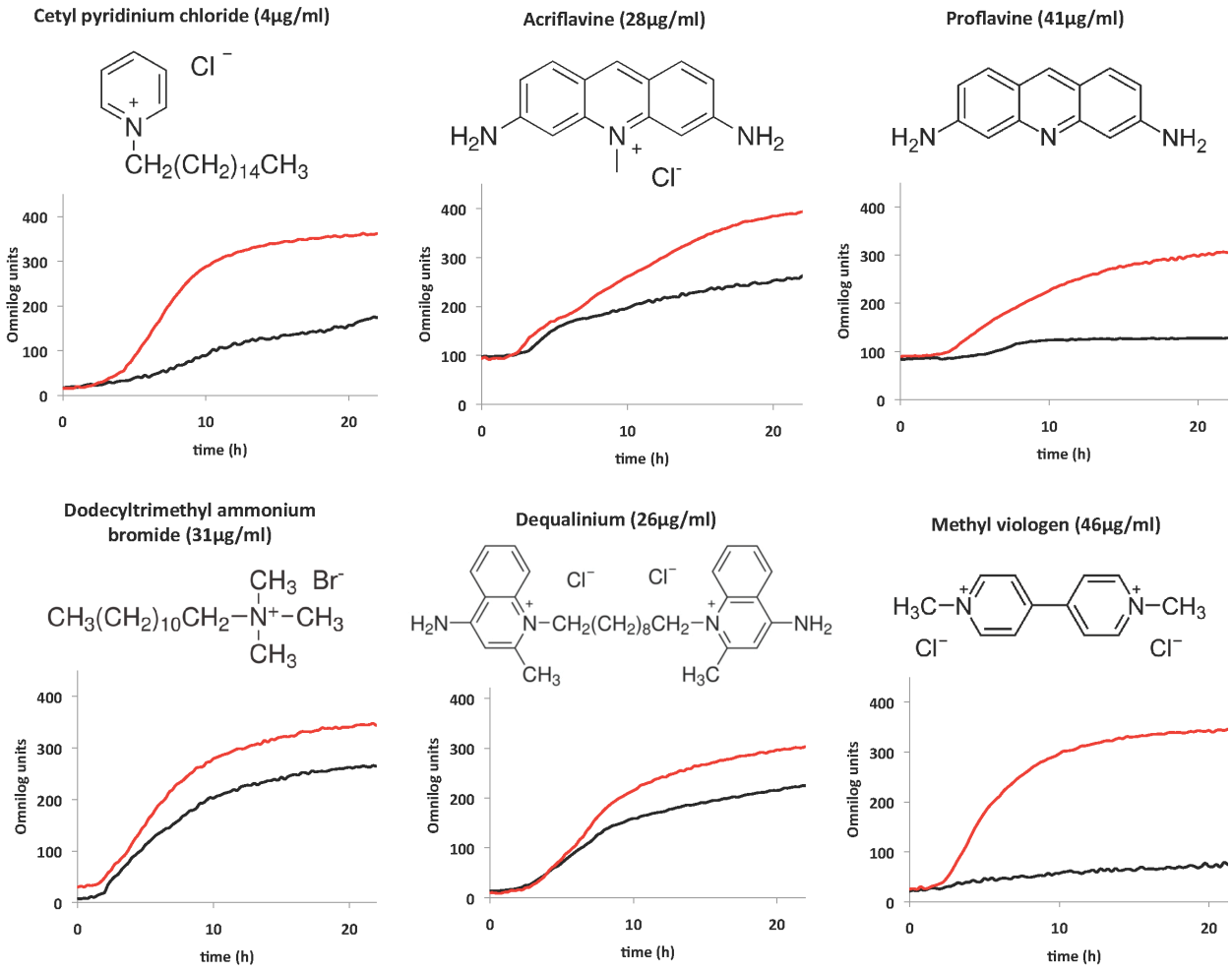
Supplementary Figures



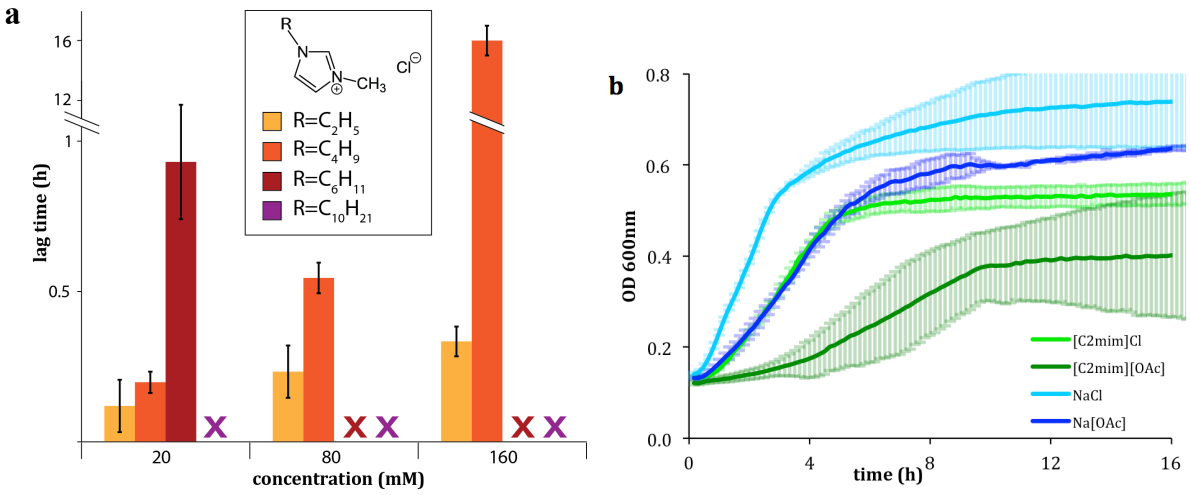
Supplementary Figure 1 | Lag time as an indication of tolerance to low levels of ionic liquids (ILs). *E. coli* harboring a plasmid with the 2,627-bp common region from the *E. lignolyticus* genome (green), or the same plasmid without the insert (blue), were cultured in media containing relatively low concentrations of [C_2mim]Cl (circles) and [C_2mim][acetate] (diamonds) for 16 h. Lag times were calculated using a curve fitting model[1]. At higher [C_2mim][OAc] concentrations, growth of the control strain was insufficient for curve fitting. Each data point represents the mean of biological duplicate measurements.



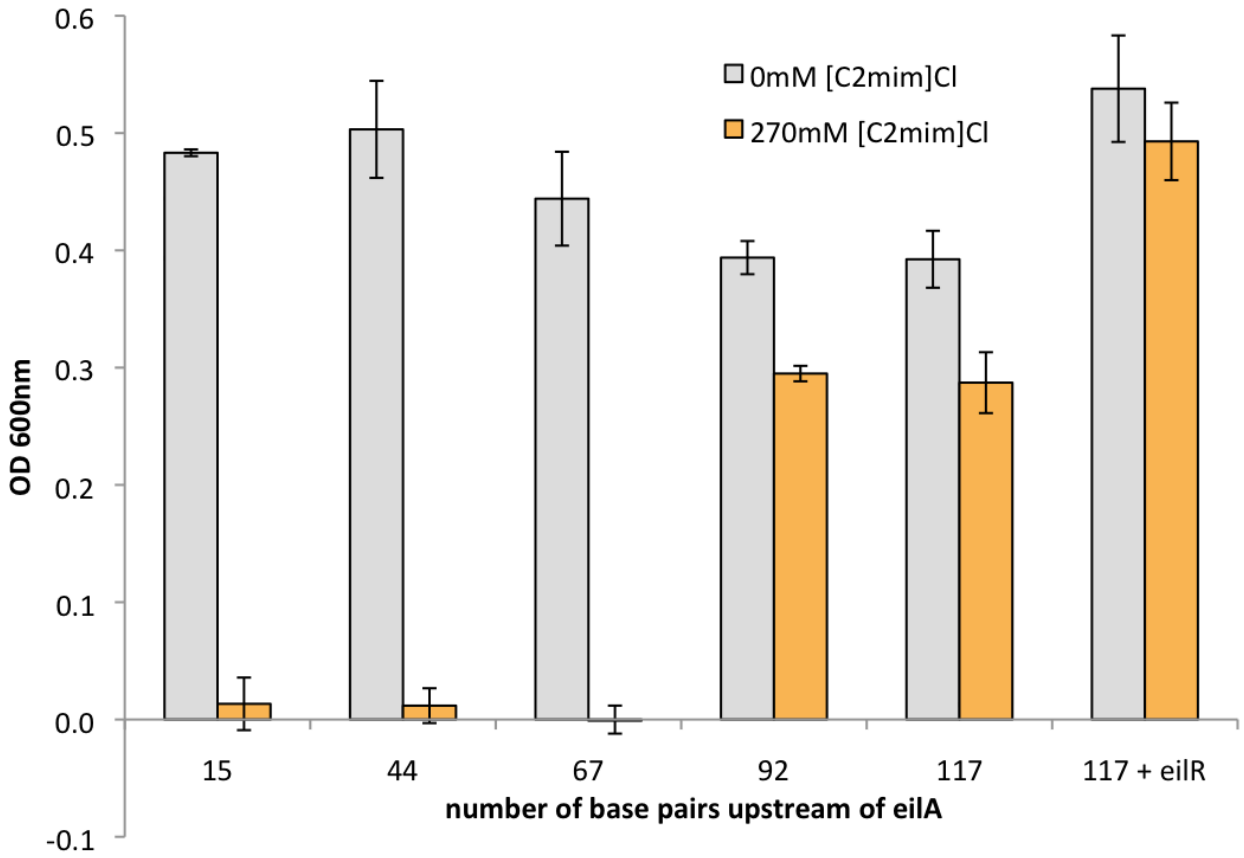
Supplementary Figure 2 | Stability of [C_2mim]Cl during growth assays. *E. coli* harboring an *eilAR* plasmid were grown in four [C_2mim]Cl concentrations for 20 hr. $[C_2mim]^+$ concentration in media before (blue diamonds) and after growth in duplicates (red circles) was measured by mass spectrometry. The "used media" values and error bars represent the means and standard deviation of biological duplicate measurements.



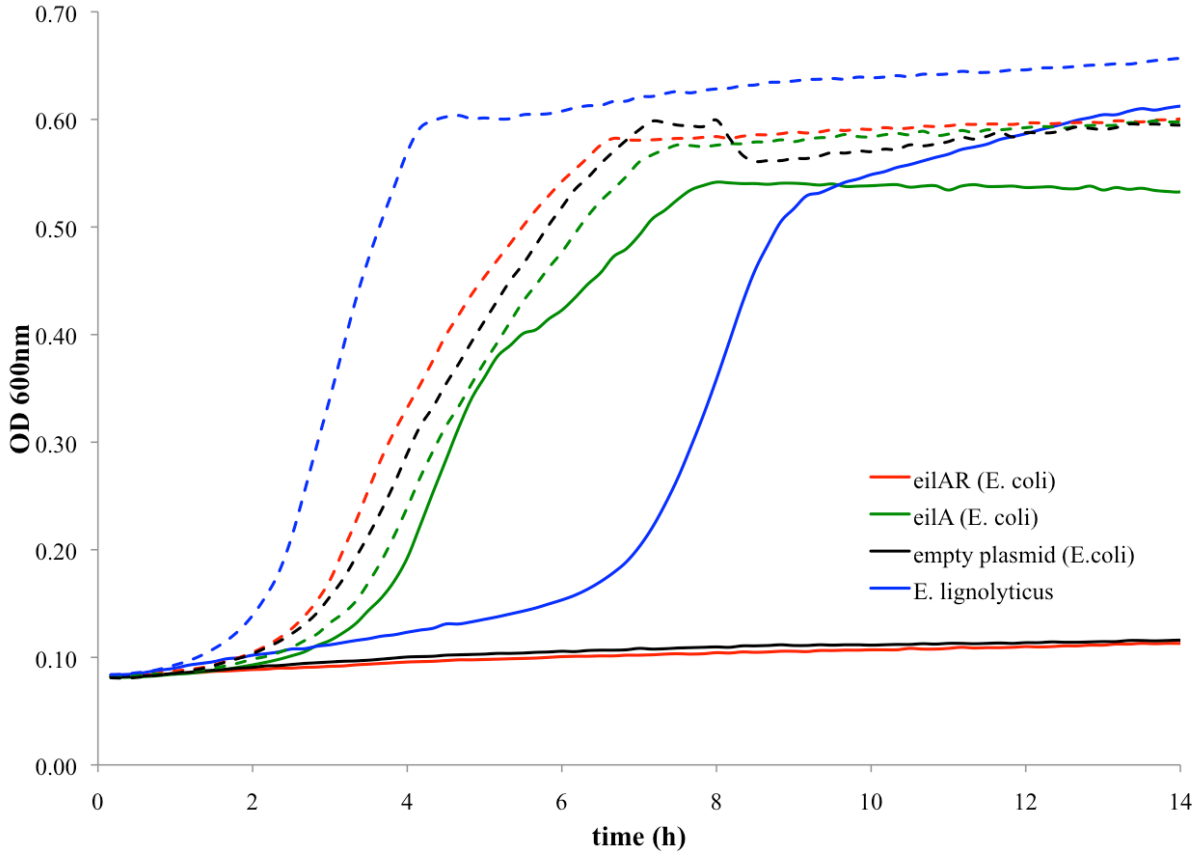
Supplementary Figure 3 | *eilA* confers tolerance to other hydrophobic cationic ammonium compounds. Using the Omnilog phenotype microarray to screen an inhibitor library, six compounds representative of the hydrophobic cationic ammonium class were detected as potential *eilA* substrates. The respiratory response in Omnilog units, used here as a proxy for cellular growth[2], was recorded in E. coli cells harboring either *eilA* and its native IG-promoter (red) or an empty plasmid (black). The curves represent the mean values of biological triplicate measurements taken every 15 min.



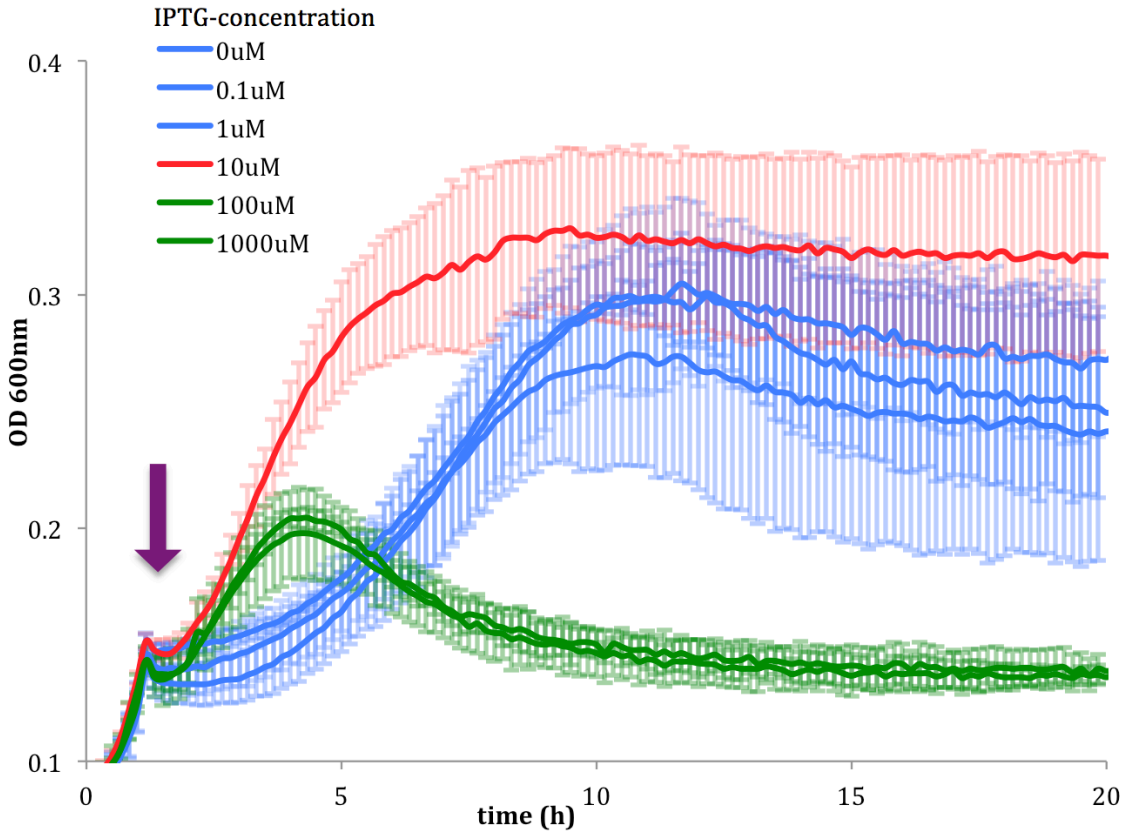
Supplementary Figure 4 | *eilA* confers tolerance to imidazolium-based ILs and salts. a. Growth of *E. coli* expressing *eilA* in media containing imidazolium chloride ILs, differing only in side chain length (inset), indicated longer alkyl groups are more toxic[3]. Lag times were calculated using a curve fitting model[1], and are presented after subtracting the lag time for a culture grown in the absence of ILs. Concentrations of ILs were as indicated; X indicates that no growth was observed. **b**. Different combinations of cations and anions confirm that $[C_2mim]^+$ and $acetate^-$ have synergistic inhibitory properties[4]. Concentrations of ILs and salts were 270mM (equivalent to 4% $[C_2mim]Cl$). For both **a** and **b**, values and error bars represent the means and standard deviation of biological triplicate measurements.



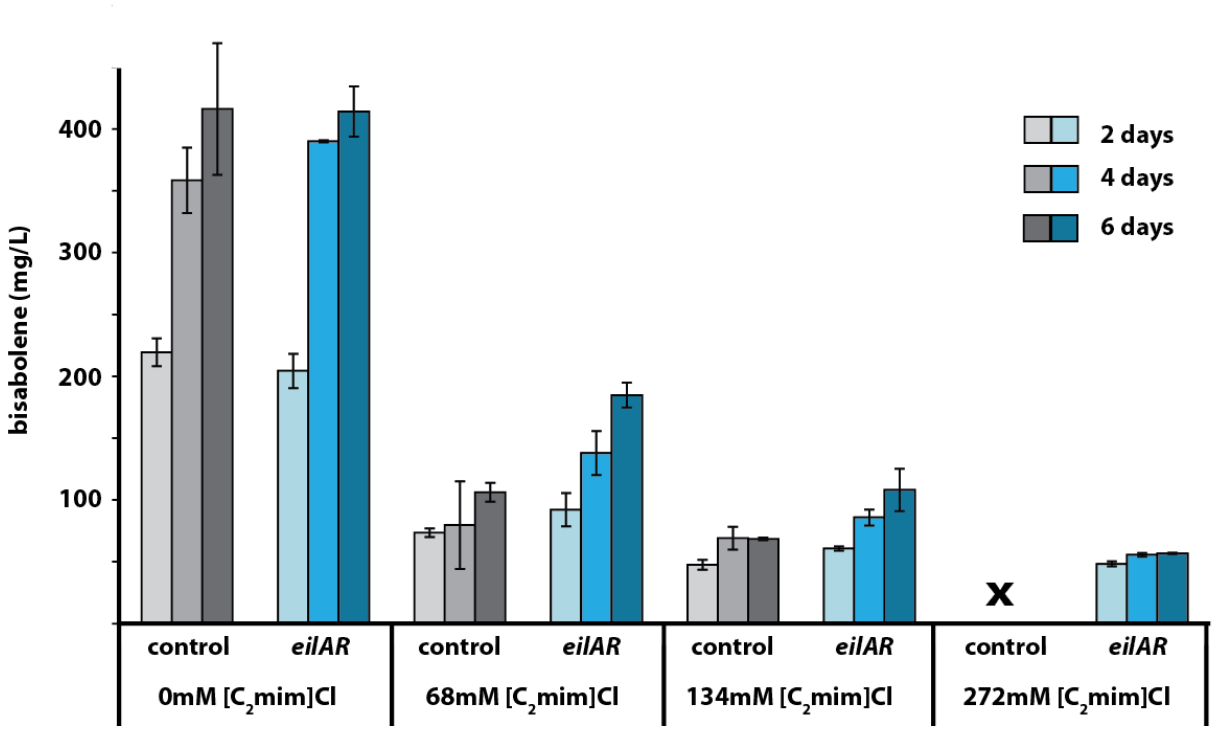
Supplementary Figure 5 | Truncation of the intergenic region (*IG*) delineates the *eilA* promoter site. Cell densities were measured in stationary phase cultures of *E. coli* harboring a plasmid containing *eilA* with truncated portions of the 117-bp *IG*. The promoter sequence does not extend further than the first 92-bp upstream of *eilA*, which was sufficient to establish [C_2mim]Cl tolerance. Control medium, gray bars; IL medium, orange bars. Values and error bars represent the means and standard deviation of biological triplicate measurements.



Supplementary Figure 6 | Resistance to the herbicide methyl viologen is conferred by the presence of *eilA* but greatly reduced by *eilR*. Cultures were grown in medium with 0.1mM methyl viologen (full lines) and without methyl viologen (dashed lines). Cells used were either *E. lignolyticus* (blue), or *E. coli* containing the following plasmids: *eilA* and its native IG-promoter (green), the entire *eilAR* cassette (red), or the control plasmid without insert (black). Curves indicate the mean values of biological triplicate measurements.



Supplementary Figure 7 | Overexpression of *eilA* in *E. coli* indicates a narrow range of tolerance to the EilA efflux pump. IPTG concentrations over a range of 0.1µM to 1.0mM were used to induce *eilA* expression from the *lacUV5* promoter in the standard growth assay. [C_2mim]Cl was added to 410mM (6% w/v) after 1 h (purple arrow). Optimum growth was achieved at 10µM IPTG, whereas severely decreased cell viability resulted from levels below and especially above this concentration, presumably as a consequence of producing too little or too much EilA efflux pump. Without IPTG, growth and IL tolerance were comparable to the lowest levels of IPTG, most likely due to basal expression of *eilA* from *lacUV5*. Curves indicate the mean values, and error bars the standard deviation, of biological triplicate measurements.



Supplementary Figure 8 | Complete bisabolene production data. *E. coli* harboring a plasmid containing the bisabolene pathway and a second plasmid either with the *eilAR* cassette (blue bars) or lacking an insert as control (gray bars). Bisabolene concentration (mg/L) was measured 2, 4 and 6 days after induction of cultures grown in three [C_2mim]Cl concentrations. X indicates that no production was observed. Values and error bars represent the means and standard deviation of biological triplicate measurements.

Supplementary Tables

Supplementary Table 1 | Plasmids constructed for IL tolerance experiments presented in this work.

plasmid ID	insert	primers (template)	replication origin	resistance	reference	restriction enzymes (backbone)
pBbS0c			SC101	Cam	[5]	
pBbS0a			SC101	Amp	[5]	
pBbE0a			ColE1	Amp	[5]	
p_sfGFP			f 1	Kan	[6]	
pTR_1	2627bp common sequence	1,2 (fosmid #3 [a])	SC101	Cam	this study	*Bgl*II, *Xho*I (pBbS0c)
pTR_2	IG-eilA	4,5 (fosmid #3 [a])	SC101	Cam	this study	*Bgl*II, *Xho*I (pBbS0c)
pTR_3	eilR-IG-eilA	3,5 (fosmid #3 [a])	SC101	Cam	this study	*Bgl*II, *Xho*I (pBbS0c)
pTR_4	eilR-IG-eilAgfp	6,7 (pTR_3) and 8,9 (p_sfGFP)	SC101	Cam	this study	none [b]
pTR_5	IG-eilAgfp	4,10 (pTR_4)	SC101	Cam	this study	*Bgl*II, *Bam*HI (pBbS0c)
pTR_6	IG-eilA	-	ColE1	Amp	this study	*Bgl*II, *Bam*HI (pBbE0a)
pTR_7	eilR-IG-eilA	-	ColE1	Amp	this study	*Bgl*II, *Bam*HI (pBbE0a)
pTR_8	eilR-IG-eilA	-	SC101	Amp	this study	*Bgl*II, *Bam*HI (pBbS0a)
pTR_9	eilA and 92bp upstream	11,12 (pTR_3)	SC101	Cam	this study	none [c]
pTR_10	eilA and 59bp upstream	11,13 (pTR_3)	SC101	Cam	this study	none [c]
pTR_11	eilA and 44bp upstream	11,14 (pTR_3)	SC101	Cam	this study	none [c]
pTR_12	eilA and 15bp upstream	11,15 (pTR_3)	SC101	Cam	this study	none [c]

a This is one of the fosmids selected in the IL screening assay, as described in the main text. It contains all *E. lignolyticus* DNA relevant in this study.

b The plasmid was constructed by Circular Polymerase Extension Cloning (CPEC).

c These PCR products include the vector backbone. They were phosphorylated prior to blunt-end self-ligation.

Supplementary Table 2 | Primers used in this study.

primer no and direction [a]		primer sequence (5' - 3') underlined bases indicate used restriction sites
1	f	CGCGAATTCAAAAGATCTCATCTCGCGGATCACGCGGATAAACGC
2	r	TTACTCGAGTTTGGATCCCTTATCCGCGGCGTCAGCGA
3	f	TCCGAATTCAAAAGATCTTTACGAAAATAACTCAAGCTG
4	f	TCCGAATTCAAAAGATCTATCTCTCTTCCCTGG
5	r	CTTACTCGAGTTTGGATCCTCACGCG
6	f	CATGGATGAGCTCTACAAATAAGGATCCTAACTCGAGTAAGGATCTCCAGG
7	r	CAGCAGAGCCAGCCGCGGTCTGGCGCACCTTTGC
8	f	GCGCCAGACCGCGGCTGGCTCTGCTGCAGGTTC
9	r	CGAGTTAGGATCCTTATTTGTAGAGCTCATCCATGCCATGTGTAATCC
10	r	TACTCGAGTTTGGATCCTTATTATTTGTAGAGCTCATC
11	r	GGCGTATTTTTTGAGTTATCGAGATTTTCAGGA
12	f	GCGATAATAACAAAAAGCTGGACAAGTGTTC
13	f	GTGTTCAACTTTCCCCCACGATC
14	f	GCAAACTGGACGGATGTCCAGC
15	f	TTATGAGGGAGATGTATGTTTCGCCAATG

[a] f, forward; r, reverse. Underlined bases indicate restriction sites used.

Supplementary References

1. Baranyi, J. & Roberts, T.A. A dynamic approach to predicting bacterial growth in food. *Int J Food Microbiol* **23**, 277-294 (1994).
2. Bochner, B.R. New technologies to assess genotype-phenotype relationships. *Nat Rev Genet* **4**, 309-314 (2003).
3. Ranke, J. et al. Lipophilicity parameters for ionic liquid cations and their correlation to in vitro cytotoxicity. *Ecotox Environ Safe* **67**, 430-438 (2007).
4. Ouellet, M. et al. Impact of ionic liquid pretreated plant biomass on Saccharomyces cerevisiae growth and biofuel production. *Green Chem* **13**, 2743-2749 (2011).
5. Lee, T.S. et al. BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J Biol Engineering* **5**, 12 (2011).
6. Pedelacq, J.D., Cabantous, S., Tran, T., Terwilliger, T.C. & Waldo, G.S. Engineering and characterization of a superfolder green fluorescent protein. *Nat Biotechnol* **24**, 79-88 (2006).

1.2 Benefits and challenges of functional genetic screening

The functional screening approach described in this chapter has proven to be a useful means to detect the targeted IL tolerance phenotype.

Sequence based approaches are another way to identify genes encoding a target phenotype, with ever-increasing sequencing data of single organisms and of metagenomic origin providing a sheer unlimited resource of genetic information. Such approaches are limited by their dependence on preceding knowledge of sequencing data, and the associated reliance on predicted annotations of homologous genes creates a bias. The accuracy of the genome annotations is well known to be inconclusive in many cases (Khatri *et al.*, 2012) because the functions of the homologous genes have not been experimentally validated .

Transcriptomics and proteomics are powerful comparative approaches to examine the global differential response of an organism in varying conditions. Candidate up- or downregulated genes can then be overexpressed or deleted to confirm their role in establishing the phenotype. Such -Omic approaches often identify gene products that are rather related to cellular stresses and basal metabolism than to the factors responsible for the targeted phenotype (Yoon *et al.*, 2011). However, using the appropriate controls, genes that are differentially expressed in response to a change in conditions can be determined and lead to testable hypotheses (Khudyakov *et al.*, 2012).

A functional genetic screening approach, like the one I applied, faces numerous challenges that need to be mastered for the detection of the genes encoding the targeted phenotype. A critical requirement is that the heterologous DNA is transcribed in the host. For this, the host RNA polymerase has to recognize the promoter of the heterologous DNA. When screening DNA from a single organism, it therefore makes sense to select a phylogenetically close screening host to increase the chance of promoter recognition. The findings presented in this chapter resulted from functional screening of *E. lignolyticus* DNA in the closely related *E. coli* host, both being members of the Enterobacteriaceae family. Transcription of the *eilA* efflux pump gene is facilitated, since its promoter resembles the characteristic architecture of *E. coli* promoters for housekeeping genes, which are recognized by the σ^{70} sigma factor (see page 72 in Chapter 2).

A more host independent transcription could be achieved with the use of a phage RNA polymerase that relies on its own vector-borne promoter. The highly processive RNA polymerase from coliphage T7 reads through host transcription termination

signals to produce unusually long transcripts of up to 20kb (Arvani *et al.*, 2012; Loeschcke *et al.*, 2013; Terrón-González *et al.*, 2013).

Another challenge in functional DNA screening is that transcription of gene(s) encoding the desired phenotype is often highly regulated and expression in the native host is induced only under certain conditions. For example, identification of the EilA pump via the phenotypic screening approach was possible because the EilR repressor responded to the pump substrates. *E. coli* expressing *eilA* constitutively is tolerant to methyl viologen but does not grow in the presence of the *eilR* repressor gene, indicating that EilR is not (sufficiently) induced by this toxic EilA substrate. This illustrates that candidate genes could be overlooked when expression is restricted by substrate-specific transcriptional regulation. Streptomyces species are hosts of numerous pathways of secondary metabolites that are of interest for drug development (Chen *et al.*, 2010; Weber *et al.*, 2003). Since a majority of such metabolites naturally functions as inhibitors against competing organisms or as signaling molecules, their biosynthetic pathways are highly regulated and expressed only under certain conditions (Bibb, 2005; Chen *et al.*, 2010). Condition-independent transcription of these typically large operons could be achieved by an RNA polymerase that does not rely on the preceding regulated bacterial promoters (as mentioned above).

In a phenotypic screen containing the entire gene pool, the chances of expression are limited by many of the reasons explained above – a “hit”, however, provides a more definite answer of being functional and contributing to the phenotype.

Chapter 2

Characterization and insights of EilR-mediated regulation

Image: cationic dyes that bind to EilR

2.1 Summary

Further investigation into the *eilAR* sequence resulted in detection of the operator site, *eilO*, and this in turn was applied to develop a sensitive EilR-regulated promoter. The promoter regulates transcription of a reporter gene in *E. coli.* This cellular biosensor was useful for the detection of cationic dyes, such as crystal violet and acridine orange, as EilR-ligands. Some of these molecules act as highly efficient inducers, with the ability to de-repress the EilR-mediated promoter at nanomolar concentrations, about five orders of magnitude lower than the previously described ionic liquid ligands. The activity of one of these inducers, leucocrystal violet, is dependent on its level of oxidation, which can serve as a non-invasive approach to monitor cellular oxidative stress in real time. The development of an activity-based assay was applied to identify potential endogenous effector molecules. Finally, I show how homologous efflux pumps of related bacteria are expressed differentially and how their cognate repressor proteins respond to specific effector molecules.

2.2 Introduction

2.2.1 Transcriptional regulation of multidrug efflux in bacteria by TetR proteins

Gene expression is primarily regulated at the transcriptional level. Therefore, a large number of promoters contain binding sites for transcription factors that activate or repress promoter activity in response to a stimulus, such as a signaling molecule. This allows the organism to adapt gene expression to its metabolic state and to external signals, such as the response to environmental stressors by defense or tolerance mechanisms such as drug efflux.

Transcriptional regulators belonging to the TetR family, named after its founding member, TetR (Hillen *et al.*, 1983) are widely distributed in bacteria and archea and mainly act as repressors. The majority of TetR encoding genes are transcribed divergently from the gene(s) they regulate and are separated by an intergenic region that is usually smaller than 200 base pairs (Ahn *et al.*, 2012). TetR regulators act as dimers and are two-domain proteins with a signal-receiving ligand-binding domain and a signal-transducing DNA-binding domain (Ramos 2005). Assignment to the TetR family is largely based on the highly conserved N-terminal DNA-binding domains, which forms a helix-turn-helix motif. In contrast, the C-terminal ligand binding domains, which typically interact with small molecules, are highly diverse, corresponding to the ability of TetR proteins to respond to a wide range of stimuli (Ahn *et al.*, 2012; Ramos *et al.*, 2005; Ulrich *et al.*, 2005). A major task of TetR family

repressors, is the regulation of drug efflux pumps. Both the repressor and the cognate transporter can have a narrow spectrum of substrate recognition and perform a specific task, such as the export of exogenous tetracyclines (Hillen & Berens, 1994) or endogenously produced antibiotics like actinorhodin in *Streptomyces coelicolor* (Tahlan *et al.*, 2007). In contrast, there are numerous examples of TetR repressors that regulate Mdr efflux pumps, such as QacR in *Staphylococcus aureus* (S Grkovic *et al.*, 1998) or EilR in *Enterobacter lignolyticus* that is described in this work. Multispecificity to drugs by the repressors must be attuned to their partner transporters in order to provide a functional regulated efflux system. The ability to bind dissimilar ligands requires a range of structural features that are often shared in multidrug binding pumps and regulators.

2.2.2 Structural mechanism of multidrug binding

A common feature in multidrug regulators and transporters is the presence of a multifaceted ligand-binding pocket (Schumacher *et al.*, 2001). The hydrophobic ligands penetrate into this pocket, where they interact with the surrounding hydrophobic residues. Such van der Waal's interactions are a major determinants to facilitate binding, since ligands prefer to bury themselves in the hydrophobic pocket rather than remain in solution where they have to disrupt the network of hydrogen bonds between water molecules (Neyfakh, 2002). Another important and distinguishing feature of multidrug binding proteins that recognize cationic ligands is the presence of buried negatively charged residues. In the *Staphylococcus* QacR repressor, four glutamate residues in the multidrug binding pocket collectively form a negatively charged pathway to guide the cationic ligands towards the "switch residues", which then trigger the conformational change that causes induction (Peters *et al.*, 2011). The hydrophobic environment of the binding site makes electrostatic attraction especially powerful, as it is not shielded by water dipoles. This stabilizing electrostatic interaction does not require perfect alignment of the positive and negative charges (Higgins, 2007; Schumacher *et al.*, 2001), but the distance between the charges has an influence on binding affinity. This was shown for BmrR, a Bacillus regulator that shares several ligands with QacR. Here, the closer the positive charge of the cationic ligand to its single glutamate in the BmrR binding pocket, the higher the binding affinity (Zheleznova *et al.*, 1999).

Multispecificity is usually not provided by multiple binding pockets, but rather by one relatively large drug binding pocket that accommodates the entire spectrum of drug-

type ligands. The architecture of the binding pocket enables diverse ligands to fit differently into the pocket, interacting with different sets of amino acid residues (Higgins, 2007; Neyfakh, 2002). In addition, the wall of the ligand-binding pocket is flexible and can change conformation upon ligand binding, which contributes to multispecificity. In the QacR repressor, structurally dissimilar ligands interact with different residues at almost non-overlapping sites within the binding pocket and interact with different amino acid side chains (Schumacher et al., 2001; 2004). Nevertheless, flexibility is limited, explaining why addition or removal of a specific side chain to some drugs can reduce binding affinity (Higgins, 2007).

In this chapter, the primary objective is to identify binding partners of the EilR protein, namely its DNA binding site and small molecule ligands, as well as to investigate the role of EilR as transcriptional regulator.

2.3 Results & Discussion

2.3.1 EilR has a high affinity to its operator

Observations described in Chapter 1 show that the presence of the *eilR* gene enables regulated expression of the EilA efflux pump in response to the imidazolium-based IL [C_2mim]Cl. To confirm that this regulation is established by *direct* interaction of the EilR protein and the intergenic region upstream of the *eilA* gene, I performed electrophoretic mobility shift assays with the purified repressor protein and truncated versions of the native 117 base pair intergenic region, which includes the promoter for the efflux pump (Figure 21).

The results of these experiments show that EilR binds to at least one region between the *eilA* and *eilR* genes. In two steps, I could confine the EilR binding site to a region located between 11 and 44 base pairs upstream of the *eilA* start codon (Figure 4). In complex with EilR, intergenic DNA containing 92 or more base pairs showed a larger mobility shift than the shorter sequences investigated. This indicates the presence of a second EilR binding site, which does not stretch further upstream than base pair 92 but requires more than 67 base pairs upstream of *eilA* for EilR binding.

In parallel to these experiments, Pavel Novichkov [1] analyzed the intergenic region of a number of bacterial repressor-efflux pump pairs in a bioinformatic approach and identified a conserved 24-base pair motif that was represented twice on the intergenic region. In the intergenic region of *eilA-eilR*, this motif is found 20-43 base pairs upstream of *eilA* (eilO1), matching my observations of an operator site in the region of base pairs 11-44. The second site of the motif (eilO2), 56-79 base pairs upstream of *eilA.* This matches with my prediction of a second operator. The consensus operator is shown in Figure 5. Annotation of the entire intergenic region, including promoter and operator sites, is analyzed later (Figure 21).

[1] (Physical Biosciences Division, Lawrence Berkeley National Labs).

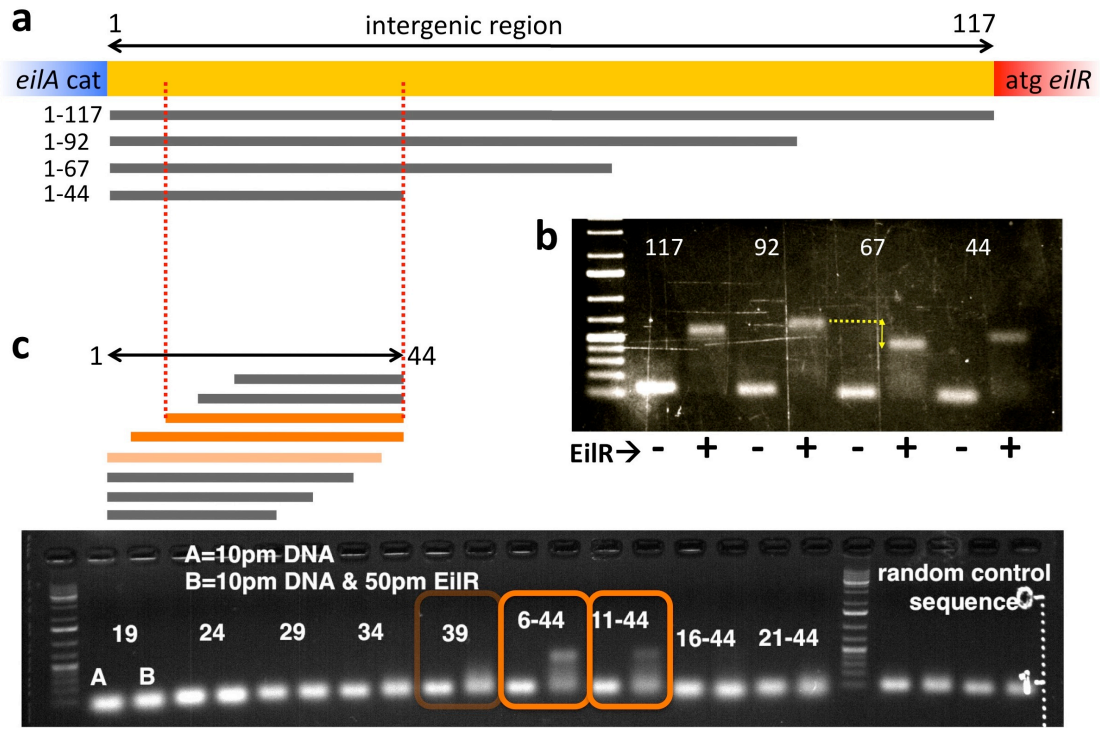


Figure 4 | Localization of one eil-operator. (a) Schematic overview of the 117bp eilA-eilR intergenic region and the four duplex oligonucleotides used in the initial round of electrophoretic mobility shift assays. Numbers represent the position of base pairs upstream of the *eilA* start codon. **(b)** The band shift between naked DNA and co-incubated EilR-DNA shows that EilR binds to all four lengths of the intergenic region. Decreased mobility of the two longer fragments indicates the presence of a second EilR-binding site (yellow arrow) further upstream of the *eilA* start codon. **(c)** The first 44bp of the intergenic region with respect to *eilA* were further fragmented in order to localize the operator site. Mobility shift of the fragments, shown in orange frames, indicates that EilR binds to a region between 11 and 44 base pairs upstream of the *eilA* start codon (red dotted lines).

Figure 5 | Motif of the 24 base pair eil-operator. This consensus sequence consists of a pair of 11 base pairs inverted repeats that are separated by two unconserved base pairs. The four strongly conserved nucleotides on each operator half-site indicate their essential role in EilR-binding.

Binding of EilR to the predicted two operators eilO1 and eilO2 was confirmed by gel mobility shift assays (Figure 6) and by in-vitro transcription/translation assays (Figure 8). The latter also indicates that [C_2mim]Cl releases EilR from its operator in a cell-

free environment at concentrations of 5 and 10mM, resulting in expression of the green fluorescent reporter protein (GFP). The decrease of GFP fluorescence observed at 20mM [C_2mim]Cl might result from the inhibition of enzymes in the delicate in-vitro reaction mix.

The results obtained in the gel mobility shift assay presented in Figure 6 show that the most conserved nucleotides are required for EilR binding. These nucleotides (four per half operator) are present in the operators of all phylogenetically related bacteria (see also Chapter 3, Figure 23). Importantly, this assay also revealed that EilR has a higher affinity to the consensus operator than to the native binding sites. Ongoing structural analysis of X-ray crystallographic data obtained from EilR complexed with half of the consensus operator [1] supports my observations on a strong and specific interaction between the repressor and its consensus DNA binding site (Figure 7). I therefore used the perfectly palindromic 24 basepairs "consensus operator" for subsequent experiments.

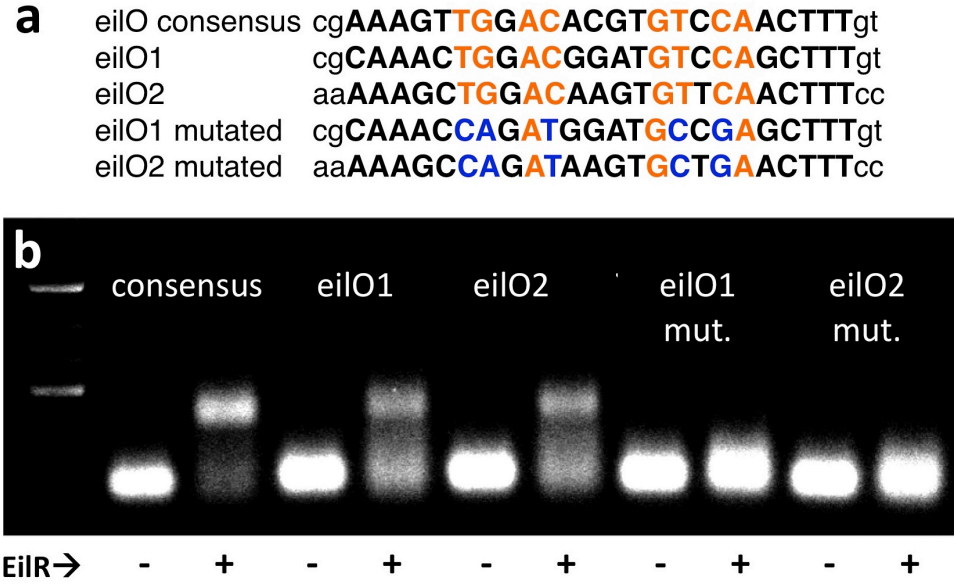


Figure 6 | The EilR repressor protein has increased affinity to a consensus eil-operator eilO. (a) Operator sequences (bold upper case) used for gel mobility shift assays. EilO1 and eilO2 are the native operator sequences located on the E. lignolyticus eilAR intergenic region. The most conserved base pairs (orange) were replaced by randomly chosen nucleotides (blue). **(b)** Electrophoretic mobility shift binding assays show specificity of EilR to its operators, eilO1 and eilO2, which is lost when conserved base pairs are mutated. The consensus operator exhibits strong affinity to EilR**.**

[1] Crystallography and structural analysis was performed by Henrique Pereira (Lawrence Berkeley National Lab.)

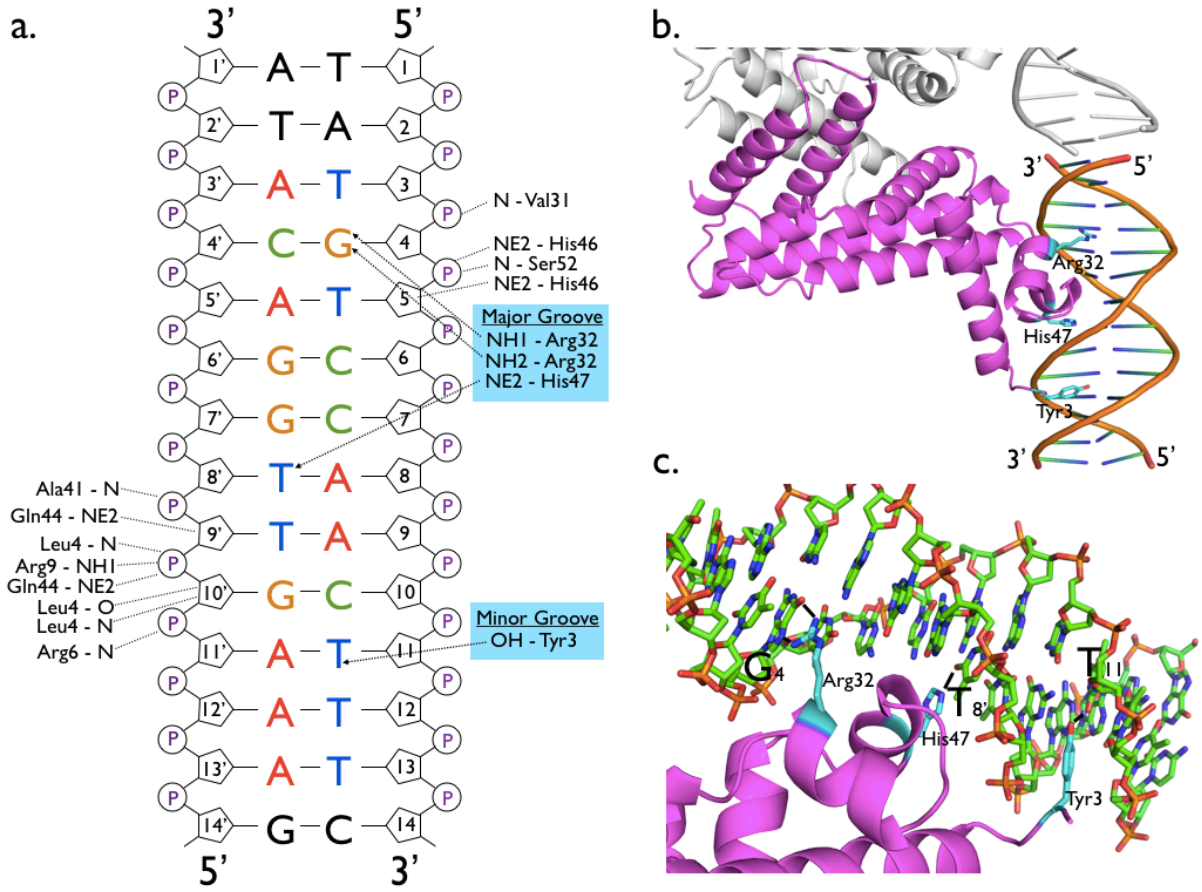


Figure 7 | Structure of the EilR-eilO complex reveals interacting residues and nucleotides. (a) Schematic representation of hydrogen bonds observed between monomer EilR and half its DNA palindromic operator sequence. The EilR transcription factor contact direct the DNA nucleotides via hydrogen bonds: Tyr3 with the DNA nucleotide $T_{11,}$ located in the minor groove region. The residues Arg32 and His47 from the Helix-Turn-Helix region of the DNA-binding domain contact nucleotides G_4 and $T_{8'}$ from the major groove region. **(b)** Overview of the EilR structure bound to half of its palindromic consensus operator DNA. **(c)** Interaction of EilR residues Tyr3, Arg32 and His47 with specific nucleotides T_{11}, G_4 and $T_{8'}$ in the DNA operator region, respectively.

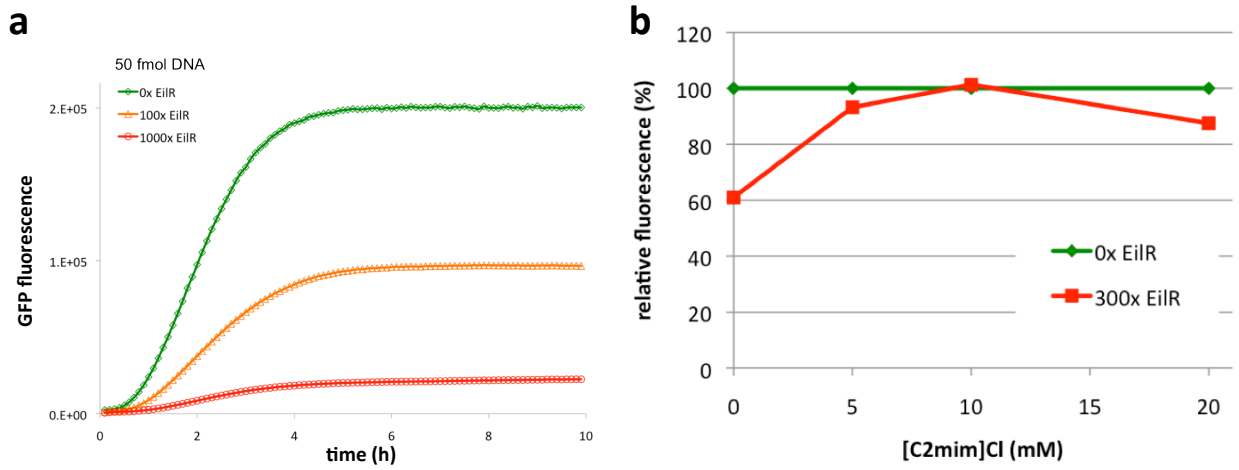


Figure 8 | Purified EilR protein binds to the intergenic region and represses in vitro gene expression. (a) On a plasmid, the intergenic region was inserted in between a T7-promoter and the ribosome-binding site of a *gfp*-gene. Fifty fmol of this DNA was added to a cell-free transcription/translation reaction with a 100- or 1000-fold molar excess of purified EilR protein (red, orange) or no EilR (green). **(b)** GFP-expression of the same plasmid as in a, in the presence (red) or absence (green) of 300x molar excess of EilR. Values are relative to GFP fluorescence of the reaction without EilR at given [C_2mim]Cl levels.

2.3.2 Development of an EilR-regulated biosensor

Using the results on EilR-DNA interactions, I built a library of randomized promoters that all contain a truncated consensus eilO operator and drive expression of the red fluorescent protein RFP as an output signal. As a result, I generated the "p1" promoter, which is regulated by EilR. This p1-promoter, in conjunction with the *rfp*-gene, was used as biosensor in *E. coli* for subsequent experiments.

It was difficult to predict how insertion of the eilO operator would affect activity of a known promoter. Randomizing the least conserved three base pairs of both the -35 and -10 sites of the consensus *E. coli* promoter increased the chance of finding a randomized sequence combination with the desired properties (Vivek Mutalik, personal communication). Selection criteria in this case included a high dynamic range (ON/OFF ratio), by a promoter that is 1) repressible by EilR and 2) has high activity in the de-repressed state. The approach taken to build this library and to detect suitable candidates is described in Figure 9. Screening of the library was performed in *E. coli* growing either in media only (repressed state) or in the presence of 300mM [C_2mim]Cl (de-repressed state). To qualify for [C_2mim]Cl tolerance, the strain independently and constitutively expressed the EilA efflux pump.

Expression of RFP turned four of the 136 screened clones visibly pink in the presence of [C_2mim]Cl, while these clones remained transparent in the absence of this IL. A clear differential RFP expression under these two conditions was also observed by fluorescence measurements. All of the selected four candidates sensitively responded to increasing concentrations of [C_2mim]Cl (Figure 10).

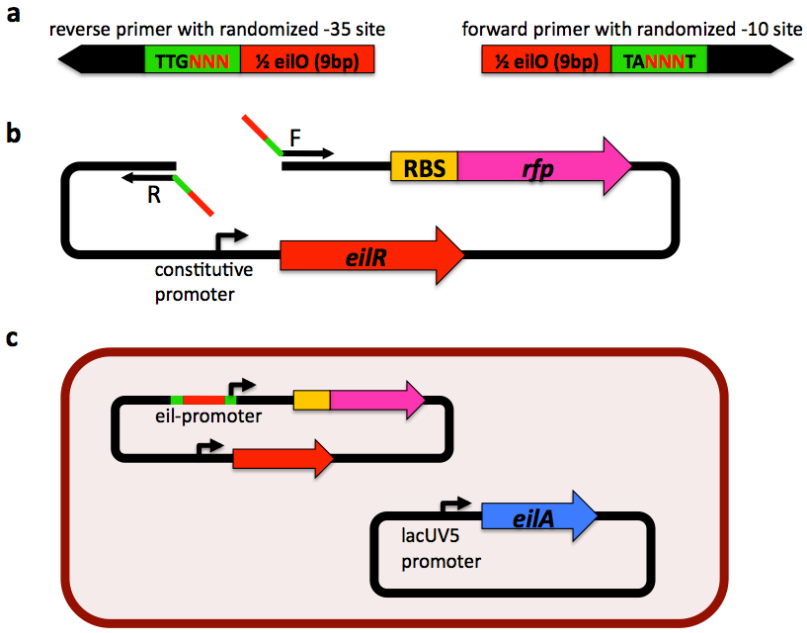
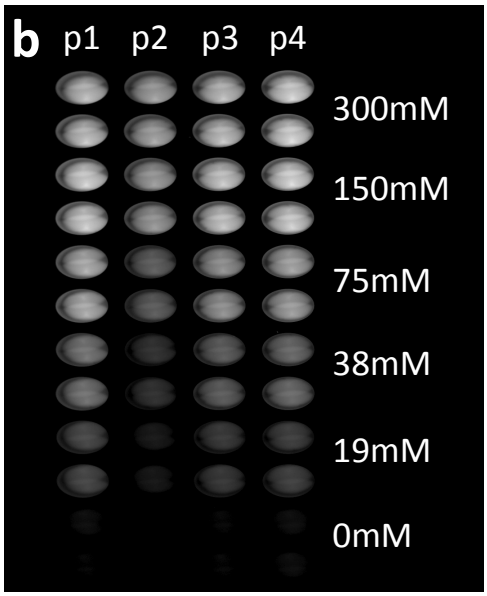
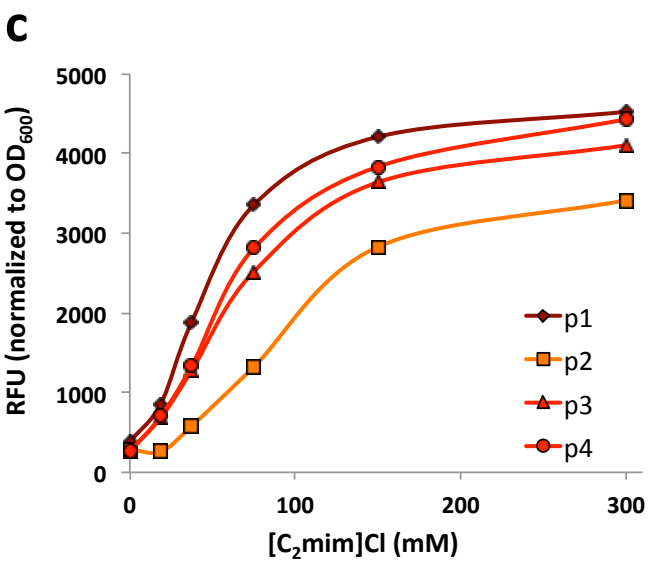


Figure 9 | Overview of the construction of an EilR-controllable reporter plasmid. (a) Oligonucleotides used to build a promoter library with randomized base pairs (N, shown in red letters) in the -35 and -10 promoter hexamer regions (shown in green). The consensus *eil*-operator (red) was truncated from 24 to 18 base pairs to fit into a typical *E. coli* spacer region between the hexamers. Each oligonucleotide contained half (9 base pairs) of the truncated eilO. **(b)** A medium copy number plasmid (p15A origin of replication) containing the *eilR* gene (red) driven from a weak constitutive promoter and an *rfp*-gene (pink) was PCR-amplified with the randomized oligonucleotides that anneal at a region upstream of the *rfp*-ribosome binding site (yellow). **(c)** The PCR products were circularized through self-ligation before being transformed into *E. coli* cells harboring a low-copy plasmid to express the EilA efflux pump driven from the independently regulated, IPTG-inducible lacUV5-promoter. EilA is needed to establish [C2mim]Cl tolerance during the subsequent screening of the library in the presence of 300mM [C_2mim]Cl. Transformed cells were grown on LB-agar plates overnight prior to screening of single colonies.

a

2.0	1.9	2.1	1.4	1.7	1.6	2.0	2.5	1.6	1.8	3.9	2.1
4.5	3.7	1.6	1.9	1.8	1.8	1.6	1.8	2.2	1.9	1.8	3.5
1.9	1.8	1.5	1.5	1.6	1.8	1.5	4.2	1.6	1.6	2.1	1.8
1.8	1.7	1.6	1.7	1.7	1.7	2.1	2.0	2.0	4.6	1.7	2.1
1.5	4.0	1.8	1.9	2.5	1.5	2.4	1.6	3.7	1.8	1.8	3.9
1.1	1.5	1.7	1.6	2.6	1.7	1.4	1.4	1.5	3.7	2.8	1.8
1.8	**11 (p1)**	1.6	1.6	3.9	1.6	1.8	1.8	6.9	2.6	1.9	2.9
1.9	4.1	1.6	1.7	1.2	1.9	2.0	1.8	2.0	3.9	1.6	2.4
1.6	1.8	1.7	1.8	1.8	1.8	1.9	**13 (p2)**	1.6	1.6	2.0	1.9
2.5	1.9	5.3	2.3	2.1	1.7	6.4	2.0	1.7	**13 (p3)**	2.1	2.0
7.7	4.1	2.2	1.9	4.7	1.8	1.9	2.2	4.8	1.9	2.1	2.0
5.0	3.0	2.6	2.5	3.4	1.8	3.9	2.6	1.8	1.8	4.3	6.1
1.8	1.8	**11 (p4)**	2.3	1.8	1.5	1.7	1.6	2.0	2.3	2.4	2.0





d

p1	TTGACAttggacacgtgtccaacTATGAT
p2	TTGACAttggacacgtgtccaacTAATGT
p3 & p4	TTGACTttggacacgtgtccaacTATGAT

Figure 10 | Selected randomized promoters containing a truncated eil-operator respond to [C_2mim]Cl. (a) Fluorescence measurements of 136 *E. coli* clones containing an *eilA*-plasmid for [C_2mim]Cl-tolerance and the *rfp*-reporter plasmid. RFP expression is driven by the randomized EilR-regulated promoters, which all contain the truncated eil-operator. Values represent the ratio of fluorescence of cells grown in 300mM [C_2mim]Cl to cells grown in media only. Clones with the highest dynamic range (ratio over 10, promoters p1-p4, marked in orange) were selected for further experiments. Note that background fluorescence was not subtracted here. **(b)** and **(c)** RFP fluorescence of cells containing promoters p1-p4 after growth at different [C_2mim]Cl concentrations was recorded in a gel-imaging station (b) or in a microtiter plate reader (c). **(d)** Sequence of promoters p1 to p4. Promoters p3 and p4 were identical. The eil-operator is marked in red, the -35 and -10 promoter boxes are shown in upper case letters.

As shown in Figure 11, the highly sensitive p1 promoter has a much stronger response to effector molecules of EilR than the native system occurring in the intergenic region. Promoter p1 was chosen for the biosensor strain. Unless otherwise stated, the biosensor used in the subsequent experiments was *E. coli* strain DH10B that contains a medium copy plasmid, from which the EilR regulated p1 promoter drives expression of the *rfp*-reporter gene. The *eilR*-gene is constitutively expressed at low levels and located on the same plasmid. Note that the additional plasmid responsible for EilA-pump expression during the screening procedure was removed from the biosensor strain.

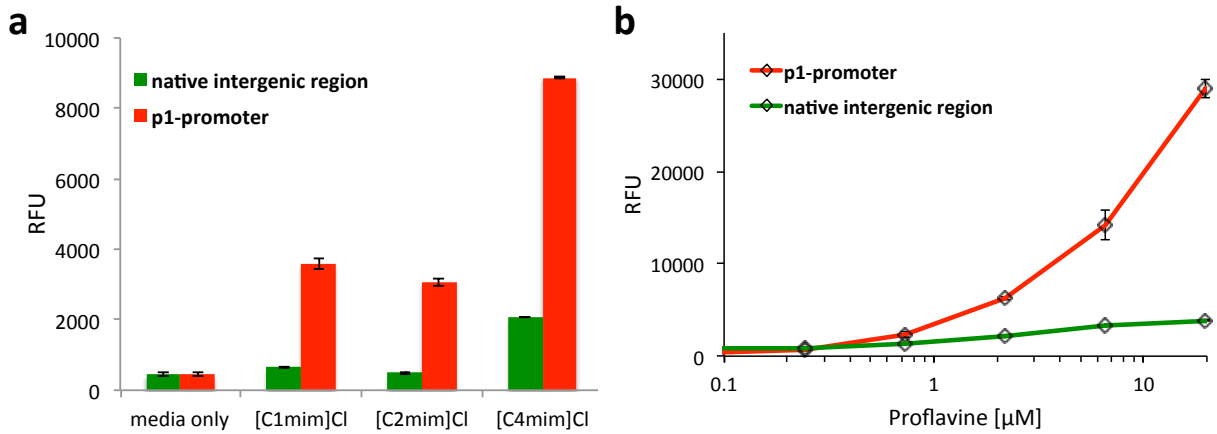
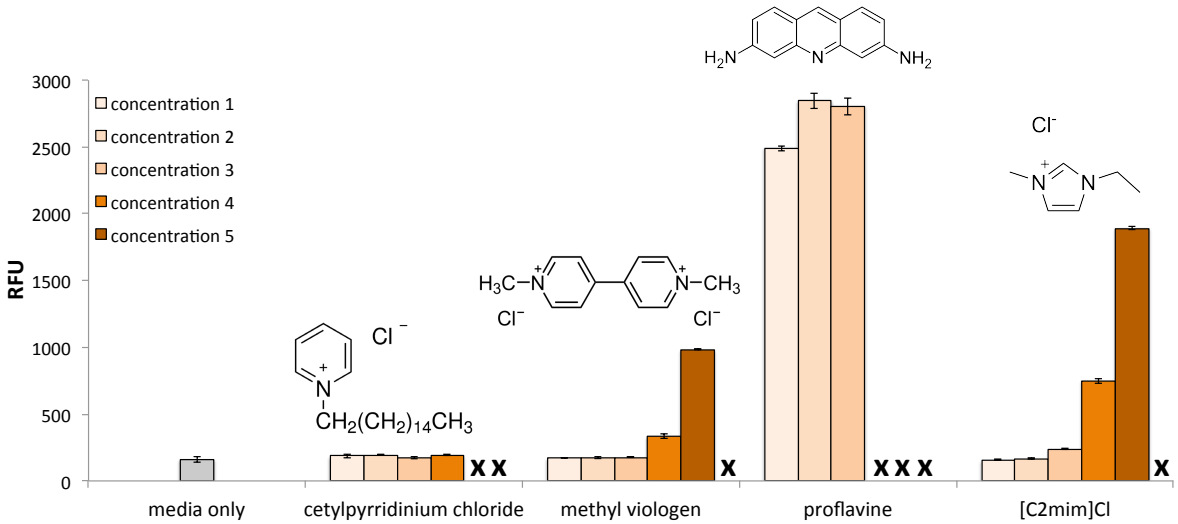


Figure 11 | The developed EilR-regulated biosensor is highly sensitive to effector molecules. Normalized fluorescence of *E. coli* expressing RFP from either the native intergenic regulatory region (green) or from the engineered p1 promoter in the presence of imidazolium-based ILs of different chain length **(a),** or in the presence of the effector molecule proflavine (see next section) **(b).** Values and error bars represent the means and standard deviation of measurements from two (a) or three (b) independently grown cultures.

2.3.3 Ligands of EilR

2.3.3.1 Cationic dyes as ligands with high binding affinity

As described in Chapter 1, I screened a library of bacterial inhibitors in *E. coli* that constitutively expresses the EilA efflux pump. EilA increased tolerance to several structurally dissimilar cationic amines (Ruegg *et al.*, 2014). Here, by exposing four of these pump substrates to the biosensor, I investigated whether they also interact with EilR. Unlike the imidazolium-based IL [C_2mim]Cl, the long-chained antiseptic cetylpyridinium chloride did not cause de-repression of the p1 promoter at all tested concentrations, of which the two highest completely inhibited growth. Methyl viologen, a bivalent cation generating reactive oxygen species, slightly induced the p1 promoter at 100μM and 200μM, but was lethal to *E. coli* at higher concentrations. Proflavine, an antiseptic fluorescent acridine dye, strongly induced the p1 promoter at a concentration of only 12.5μM, resulting in RFP levels higher than those of all other tested substrates (Figure 12).

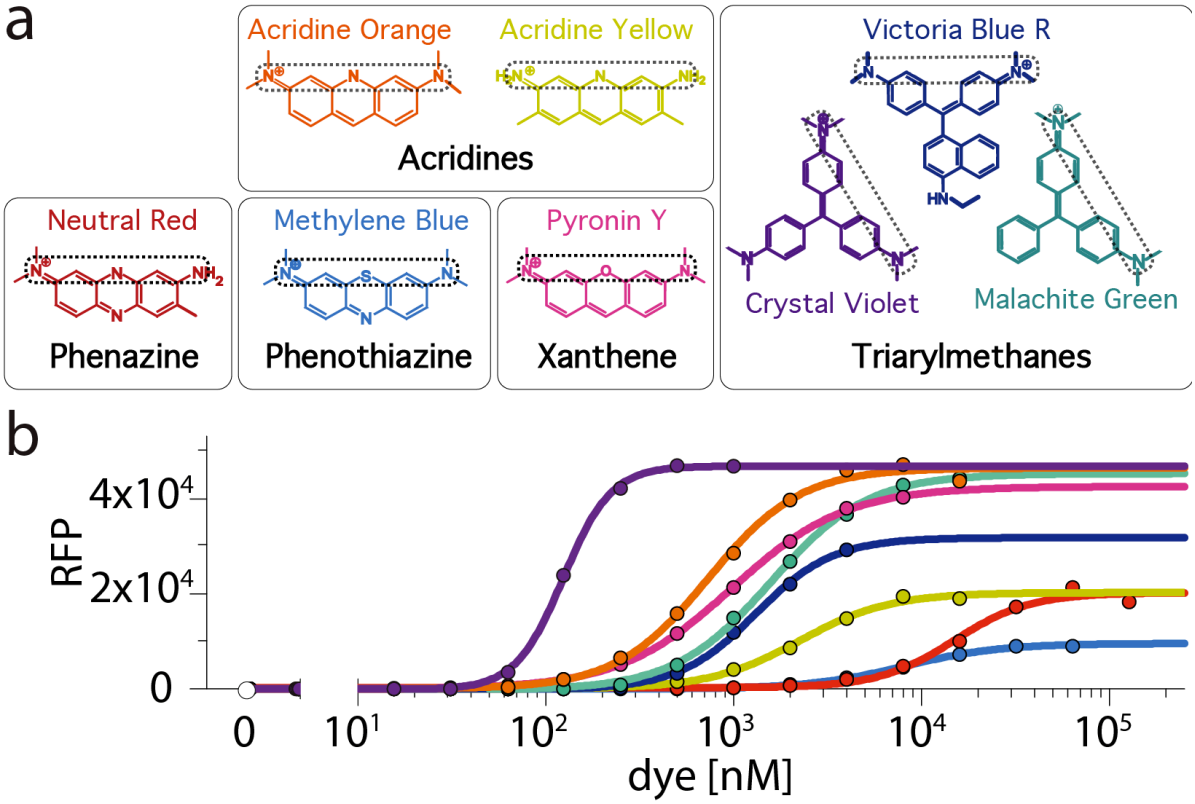


	concentration 1	concentration 2	concentration 3	concentration 4	concentration 5	concentration 6
cetylpyrridinium chloride (uM)	1.25	2.5	5	10	20	40
methyl viologen (uM)	12.5	25	50	100	200	400
proflavine (uM)	12.5	25	50	100	200	400
[C2mim]Cl (mM)	0.16	0.8	4	20	100	400

Figure 12 | Screening of EilA-pump substrates with the p1-reporter. Normalized fluorescence of stationary phase *E. coli* harboring the p1 reporter plasmid grown in different concentrations of EilA substrates. "X" indicates that no growth was observed. Values and error bars represent the means and standard deviation of measurements from two independently grown cultures.

Given the intense response of EilR to proflavine, I expanded the screen to other readily available cationic dyes belonging to the acridine, triarylmethane, phenothiazine, phenazine and xanthene families (Figure 13). As in the previously characterized ILs, all of these dyes are monovalent cations hydrophobic in character, and with a positive charge on the nitrogen. While this nitrogen atom is part of the imidazolium ring in ILs, it forms an alkylated amine group in the tested cationic dyes. The acridines, phenothiazines, phenazines and xanthenes are tricyclic, planar molecules with a heterocyclic central ring and two equally positioned amine groups. In contrast, the triarylmethanes have a propeller-like geometry, with the aryl rings projecting from the central coordination plane at an angle of 27.7° (Lovell *et al.*, 1999).

All tested dyes had an ability to de-repress the EilR-controlled biosensor strain. However, the response drastically differed between these molecules in various aspects: First of all, the concentrations needed to establish half of the maximally achievable de-repression (EC_{50}) varied, with EilR showing the most sensitive response to crystal violet – 100 times less than neutral red. In addition, the response amplitude (y_{max}) varied between inducer compounds. For example, saturating concentrations of methylene blue resulted in RFP-levels that were about five times lower than those induced by crystal violet or acridine orange. Furthermore, the "steepness" of the response curve, delineated by the Hill-coefficient, also depended on the inducer molecule: The most sensitive reaction to changing inducer concentrations was observed with crystal violet, while the response triggered by increasing concentrations of other dyes, such as acridine orange was more gradual.



	Crystal Violet	Victoria Blue R	Malachite Green	Acridine Yellow	Acridine Orange	Neutral Red	Methylene Blue	Pyronin Y
EC_{50} [nM]	125	1315	1584	2257	736	15039	8097	987
y_{max}	46990	31908	45395	20324	46600	20319	9622	42658
n	3.4	2.1	1.7	1.8	1.8	2.1	1.7	1.4

Figure 13 | Response of an EilR-regulated promoter to cationic dyes. (a) Chemical structure of cationic dyes from different compound classes that have affinity to EilR. The dotted frames indicate a pattern that these compounds have in common. **(b)** Fluorescence of stationary *E. coli* cultures expressing RFP from the EilR-regulated DE20-eilOx-promoter (see chapter 3) after growth in EZ-Rich media containing different cationic dyes in increasing concentrations. EC_{50} indicates the dye concentration at which the promoter is de-repressed by 50%. y_{max} represents the maximally achievable RFP expression at saturated dye concentrations. The Hill-coefficient (n) indicates the steepness of the response curve. Constants and response curves were generated by nonlinear regression (4-parameter variable slope model) using values of fluorescence measurements from three independent experiments. The image shows aqueous solutions of six dyes. Colors of structural formulae, curves and fields in (a) and (b) correlate and represent the color of the respective dye.

Crystal violet possessed the most optimal properties to release EilR from its operator in this in-vivo assay. Based on the response of the biosensor to a series of triarylmethanes (Figure 14), I hypothesize that at one point, the molecule becomes too large to fit into the binding pocket, preventing its proper allocation towards the residues that trigger the conformational change for induction. Thus, increasing the bulkiness of crystal violet by one (victoria blue R) or two (victoria blue B) additional aromatic rings could explain the strong decrease in potency. The same end-effect was also observed with compounds smaller than crystal violet. Malachite green differs from crystal violet only by the lower number of dimethylamine groups. These groups might serve as anchors due to steric and/or electrostatic affects, which promote "attachment" to the ligand-binding pocket. Absence of one anchor would explain the decrease in activity observed with malachite green.

The three planar, tricyclic ligands, acridine orange, pyronin Y and methylene blue share the same structural architecture, including the position of the two dimethylamine groups. Therefore, their varying property as an inducer must originate from the different atoms (N, O, S) that form the heterocyclic ring, the only difference between these molecules. Methylene blue triggers a much weaker response, both in terms of EC_{50} and maximally achievable induction (y_{max}). It is not clear how the larger sulfur atom and the presence of an additional nitrogen atom in the ring structure contribute to the lower potency and efficacy of this molecule compared to acridine orange and pyronin Y.

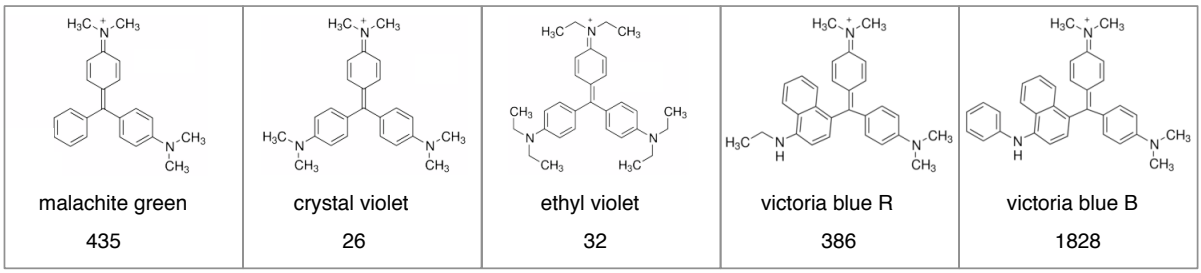


Figure 14 | Triarylmethanes and their ability to trigger an EilR-switch. Five triarylmethane compounds with increasing bulkiness are shown. Analogously to the approach illustrated in **Figure 13**, these compounds were tested for their ability to induce an EilR-regulated promoter. Numbers indicate the EC_{50} value, the concentration [nM] at which the EilR-regulated promoter is de-repressed by 50%. These values were generated by nonlinear regression (4-parameters variable slope model) using fluorescence measurements from three independent *E. coli* cultures grown to stationary phase in EZ-Rich medium containing 0.2% glucose. Here, the autoregulated ig-eilAR-v2 promoter, described in Chapter 3, was used.

It should be kept in mind that my interpretations of EilR-ligand interactions are based on observations made with the *E. coli* biosensor. Sensitivity of such an *in vivo* system to an inducer molecule is dependent on several factors: The strength of the promoter that drives expression of the repressor co-determines cellular repressor levels. At high levels, more inducer molecules are needed for de-repression, since an overrepresentation of "inactive" repressors that are not bound to an operator compete for ligand binding. In addition, host metabolism also affects response of the EilR-regulated promoter to effector molecules. Although it is unlikely that *E. coli* degrades chemical dyes, these molecules probably do not penetrate through the membrane equally well. Also, host efflux pumps like the important *E. coli* multidrug transporter AcrB might expel the inducers (Murakami *et al.*, 2004). Thus, the applied initial concentrations do not represent intracellular levels of these molecules. *In vitro* experiments, such as electrophoretic mobility shift assays with the purified EilR and dyes could serve as a complementary approach that would enable elucidation of the $K_{D,}$ the dissociation constant representing direct protein-ligand affinity.

2.3.3.2 Leucocrystal violet as agent for monitoring oxygen stress

While the cationic crystal violet de-represses the EilR-mediated biosensor at nanomolar concentrations, the repressor is relatively inert to the reduced, uncharged and colorless form called leucocrystal violet. Reactive oxygen species, such as the hydroxyl radical, have the ability to oxidize leucocrystal violet to its colored cation, crystal violet (Cohn *et al.*, 2005).

Under aerobic conditions, biological systems are exposed to oxidative stress caused by reactive oxygen species. Formation of the hydroxyl radical from oxygen or hydrogen peroxide is catalyzed, for example, by Fe^{3+} ions present in the growth media via the Haber-Weiss and the subsequent Fenton reaction (Carlioz & Touati, 1986). Thus, the radicals generated in such conditions indirectly de-repress EilR-mediated promoters by oxidizing the inert leucocrystal violet to its active form, leading to expression of the *rfp* reporter gene (Figure 15). Besides confirming that the cationic form of a molecule is crucial for EilR affinity, this signal-amplifying reaction might therefore serve as a sensitive indicator to monitor real time oxidative stress in a non-invasive way.

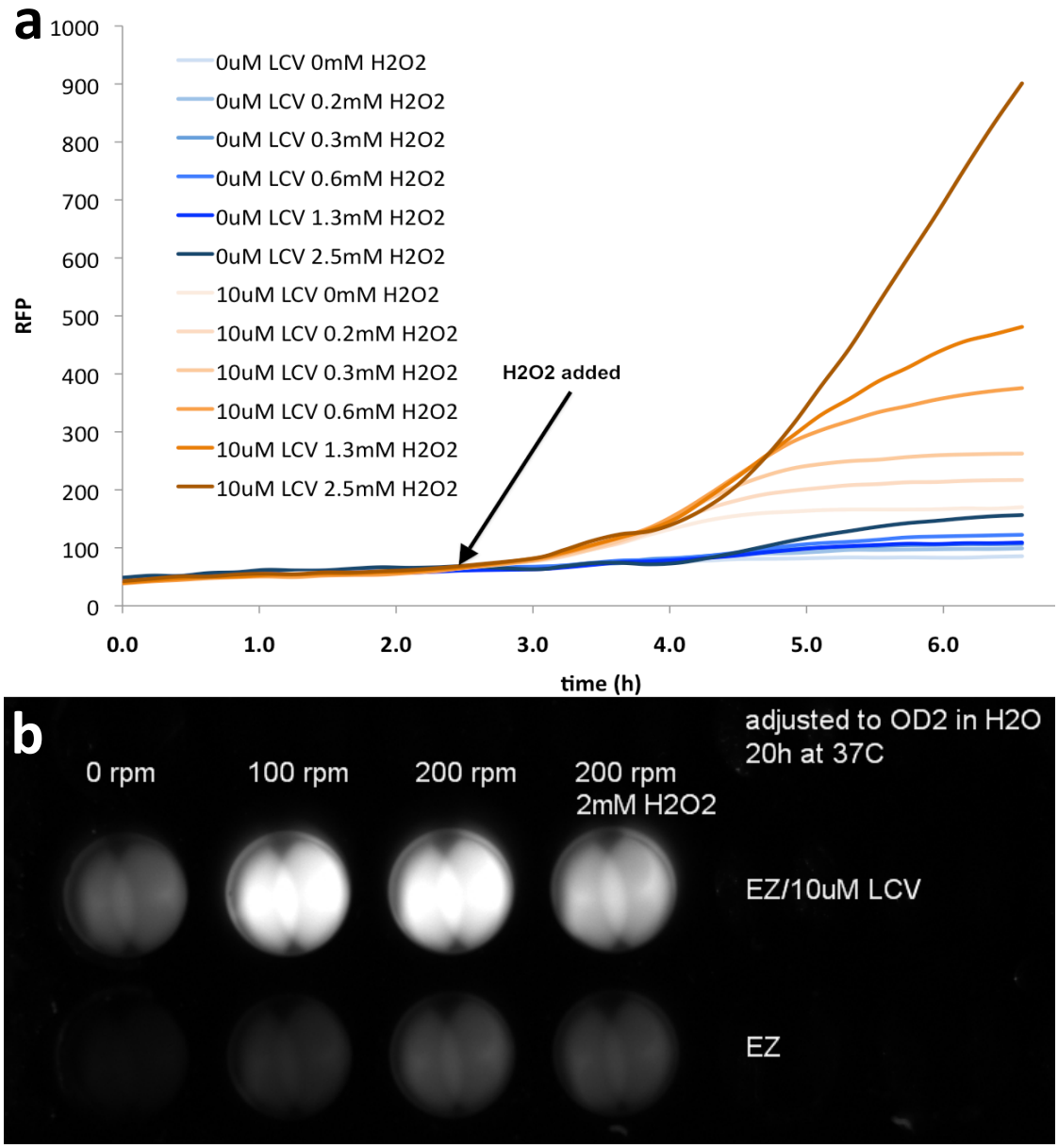


Figure 15 | Response of an EilR-regulated promoter to leucocrystal violet under different oxidizing conditions. (a) Real-time fluorescence measurements of *E. coli* cultures expressing RFP from an EilR regulated promoter. Cells were grown in the absence (blue lines) or presence (orange lines) of 10µM leucocrystal violet. Oxidative stress was induced in early exponential phase by adding different concentrations of hydrogen peroxide (H_2O_2). **(b)** RFP fluorescence of the stationary phase reporter strain after growth under different aeration conditions either with or without 10µM leucocrystal violet (LCV). Cells were grown in 5mL culture tubes shaken at 100, 200 or no rpm. In a and b, cells were grown in EZ-Rich defined media containing 0.2% glucose. Influence of the 10µM iron sulfate present in the standard media composition was not investigated. The reporter plasmid ig-eilAR-v2, which is further described in chapter 3, was used here.

2.3.3.3 Growth conditions influence biosensor activity

Here, I would like to briefly describe observations on basal activity of the biosensor in certain media types, and a possible approach that might serve for the future identification of natural effector molecules. I observed that the intensity of background RFP expression in the biosensor strain was dependent on media composition. When cells were grown in EZ-Rich defined medium with glucose or

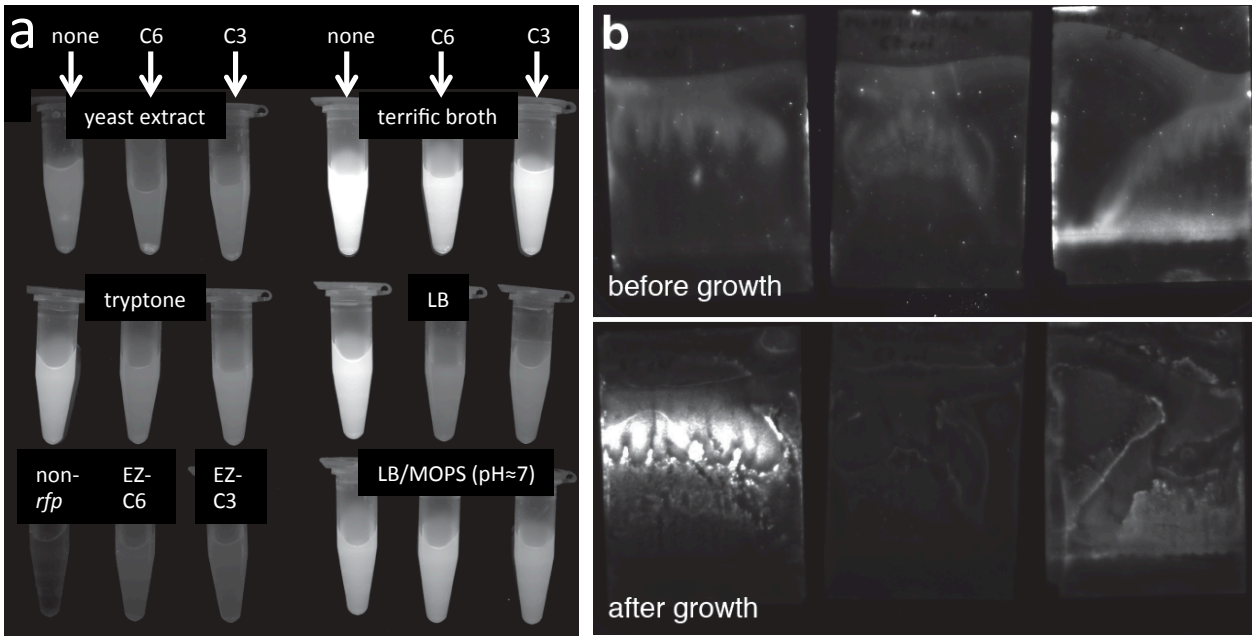


Figure 16 | Long-time exposure of the biosensor strain reveals media-dependent differential background activity of the EilR-regulated promoter. (a) The *E.coli* biosensor was grown to stationary phase in different growth media with 0.2% glucose (C6), 0.2% glycerol (C3) or without additional carbon source (none). Cells were resuspended in phosphate buffer at an equal density (OD_{600}=1) for imaging RFP fluorescence. **(b)** Fluorescence of the biosensor growing on *E. coli* cell extracts that were fractionated by thin layer chromatography (TLC). Top: TLC plates with fractionated extracts of stationary phase *E. coli* that were either grown in LB (left) or in EZ Rich/0.2% glucose medium (center). LB medium was directly applied on the right TLC plate. Bottom: The same TLC plates after incubation with the biosensor strain. This was done by submerging the dried TLC-plates in EZ Rich medium containing *E.coli* biosensor cells followed by overnight incubation.

glycerol as sole carbon source, background activity was barely detectable. In contrast, by using a complex medium such as terrific broth or LB broth, which both consist of yeast extract and peptides from trypsin-digested casein, background activity of the EilR-regulated promoter significantly increased. In LB broth, a lack of additional carbon sources, such as glucose or glycerol, resulted in a pH increase, which triggered activation of EilR-regulated promoters. The carbon source had less of an effect on promoter activity in LB broth that was buffered at pH 7 (Figure 16a).

In the activity-based approach (Figure 16b), the biosensor was exposed to cell extracts that were fractionated by thin layer chromatography. The extracts originated from cells that were grown either in LB or in defined, glucose containing media. The sensor was only activated by certain cell fractions grown in LB, but not the ones grown in a defined medium or in the presence of fresh LB.

Figure 17 illustrates that an increase in pH correlated with higher activity of a biosensor strain. RFP levels gradually rose from pH 6.8 to pH 8.5, both in the presence and absence of glycerol. A further increase of alkalinity (pH 9.5) was severely toxic to the cells, but despite very little growth, overall RFP amounts in this condition still exceeded those at all other, lower pH levels.

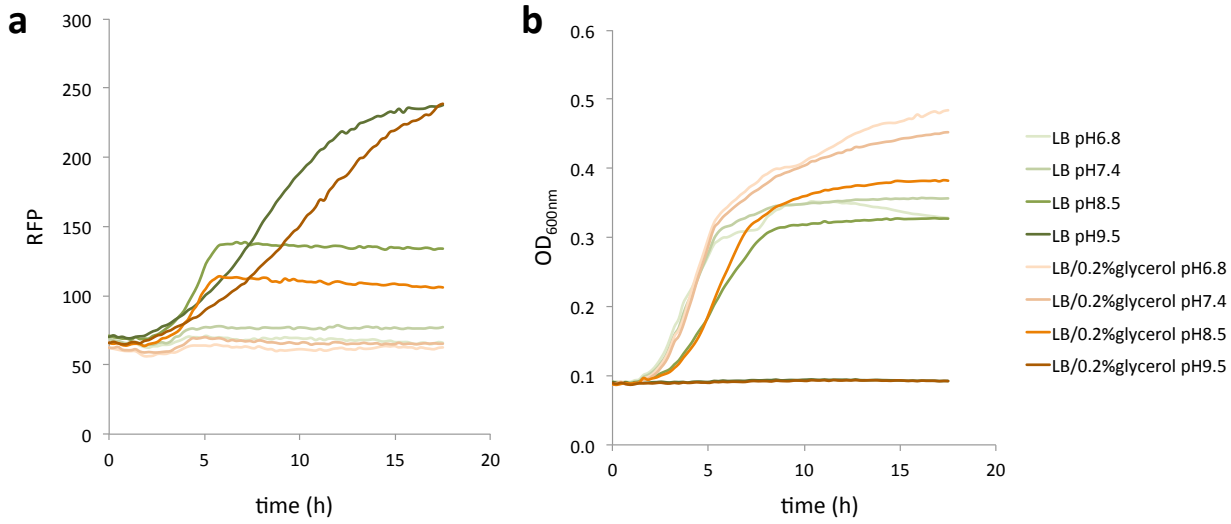


Figure 17 | Increased activity of EilR-mediated promoters correlates with increasing pH. **(a)** Total RFP fluorescence of *E. coli* containing the EilR-regulated reporter plasmid ig-eilAR-v2 (see chapter 3) growing in LB media without (orange) and with (green) addition of 0.2% glycerol, buffered at four different pH-levels. **(b)** Cell density of the same cultures, measured at 600nm. Curves represent the average values obtained from three independently grown cultures.

2.3.4 Regulation of homologous efflux mechanisms in related bacteria

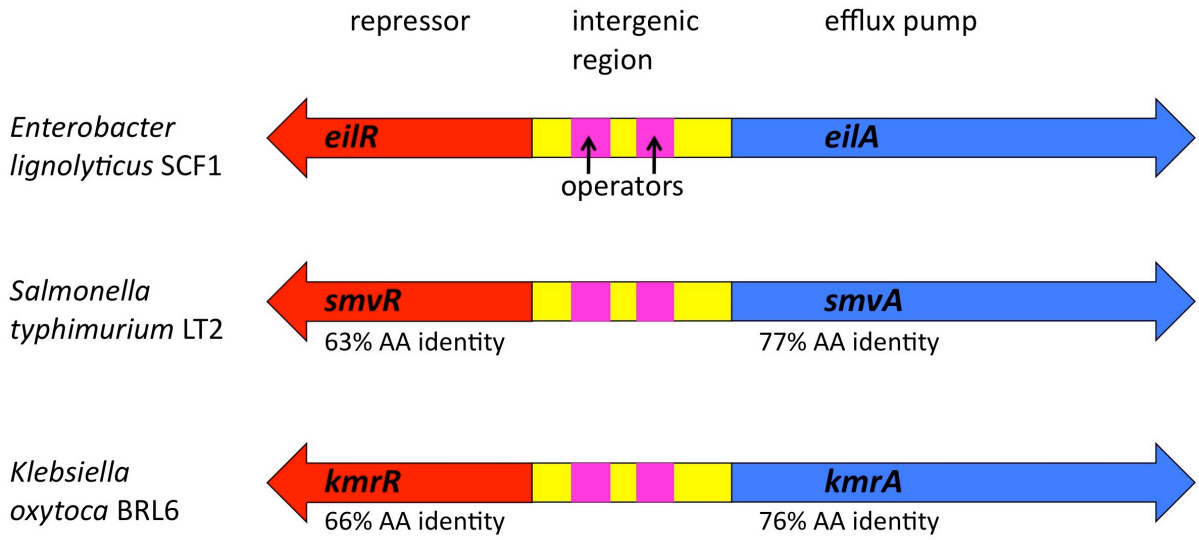


Figure 18 | Loci in *Salmonella typhimurium* and *Klebsiella* oxytoca with homology to the *E. lignolyticus eilAR* region. Annotated repressor genes (red) and the divergently transcribed efflux pump genes (blue) are separated by an intergenic region (yellow) that contains two palindromic sites, which resemble the conserved motif of the eilO consensus operator (pink). The percentages of amino acid identity with respect to the EilR and EilA proteins are indicated.

The enterobacteria *Salmonella typhimurium* LT2 (Santiviago *et al.*, 2002) and *Klebsiella oxytoca* BRL6-2 (Woo *et al.*, 2014) contain a chromosomal region that is homologous to the *Enterobacter lignolyticus eilAR* locus. As illustrated in Figure 18, an intergenic region with two palindromic sites, which are highly similar to the consensus eilO operator, separates the divergent pump and repressor genes. It is likely that this region has a similar function in all three organisms, namely by providing an autoregulated system for multidrug efflux.

2.3.4.1 The *Salmonella* SmvA pump expels [C_2mim]Cl

The SmvA efflux pump of *Salmonella typhimurium,* an EilA homologue of with 77% amino acid identity, is the major factor for tolerance to methyl viologen (Santiviago *et al.*, 2002), a bipyridinium cation that is also a substrate of EilA. It is therefore reasonable to hypothesize that SmvA also confers tolerance to the imidazolium-based IL [C_2mim]Cl. This was confirmed in growth experiments. Wild-type *S. typhimurium* is able to withstand [C_2mim]Cl concentrations of ca. 400mM (6%), while no growth was observed when levels were elevated to 540mM (Figure 19). The level

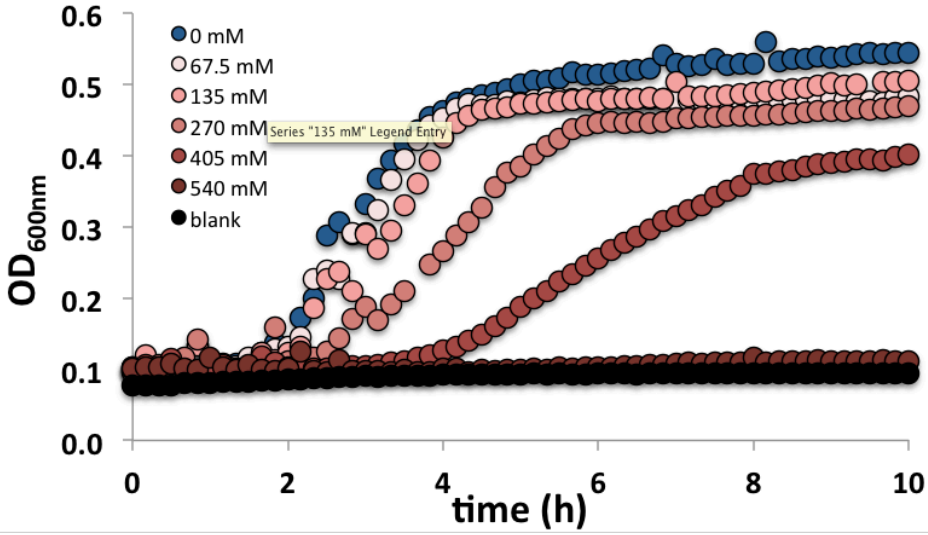


Figure 19 | Growth of *Salmonella typhimurium* LT2 in the presence of [C_2mim]Cl. Cell density of cultures growing in EZ-Rich/0.2% glucose supplemented with increasing levels of [C_2mim]Cl.

of tolerance to this IL is comparable to that of *E. lignolyticus* (Khudyakov *et al.*, 2012; Ruegg *et al.*, 2014).

In a next step, *smvA* and the upstream intergenic region were introduced into *E. coli* either with or without the divergently transcribed *smvR* repressor gene. In both cases, the SmvA pump confers [C_2mim]Cl tolerance (Figure 20). Analogously to observations with EilA (Chapter 1, Figures 2 and S7), constitutive SmvA expression, which is caused by the absence of the SmvR repressor, resulted in severely decreased growth. While weak growth was still observed when *eilA* was constitutively expressed from a high copy plasmid at high [C_2mim]Cl levels, overexpression of *smvA* from its native promoter under the same conditions completely halted cell growth. This was probably caused by a stronger activity of the native *smvA*-promoter compared with the *eilA*-promoter, resulting in high pump levels that cause cellular stress (Figure 20)(Wagner *et al.*, 2007). This observation aligns with alternate sequences of the two promoters (Figure 21): While the -35 sites are identical, the *S. typhimurium* -10 site (TAGGAT) has greater similarity to the *E. coli* (and likely *Enterobacteriaceae* in general) consensus promoter sequence (Lisser & Margalit, 1993) compared with *E. lignolyticus* (CACGAT), see also p. 72 in Chapter 3. SmvA overexpression due to stronger promoter activity halts growth in conjunction with the stress imposed by [C_2mim]Cl. Stressful pump overexpression is kept under control by the repressor. Thus, the presence of the *smvR* gene enables growth to take place all scenarios, independent of [C_2mim]Cl concentration and *smvA* copy number.

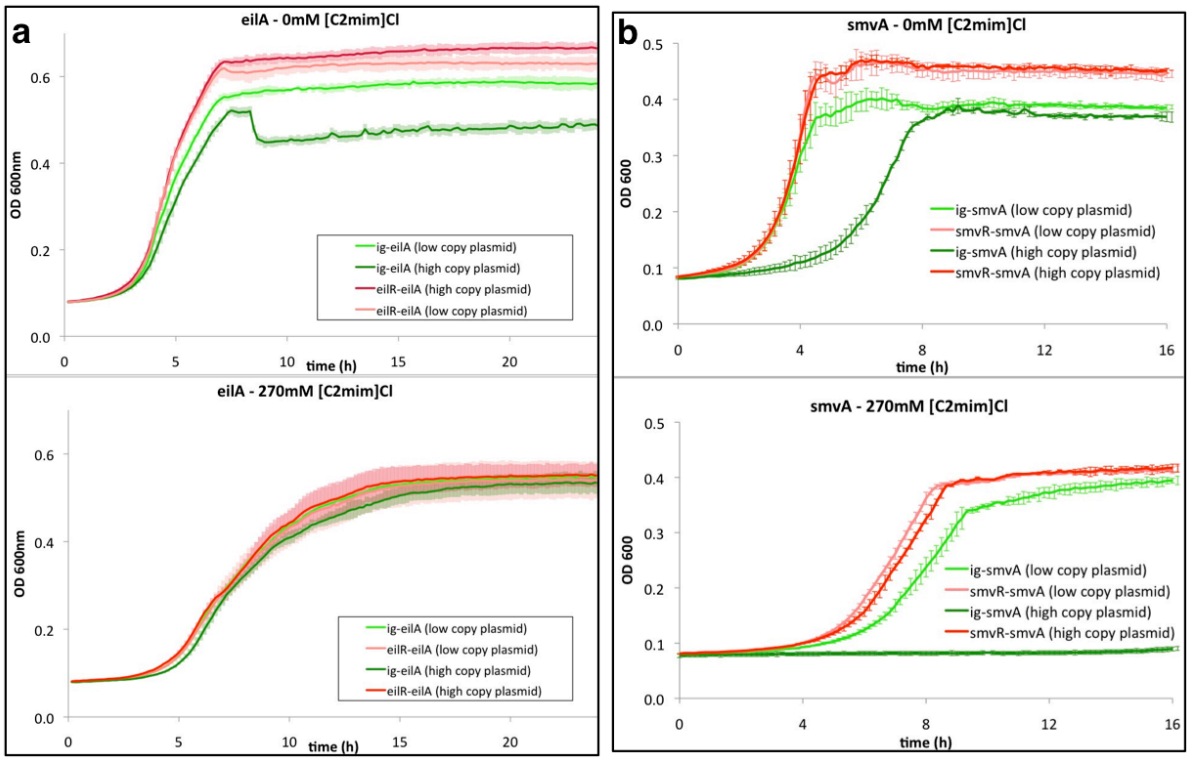


Figure 20 | Differential growth of *E. coli* expressing EilA or SmvA from their native promoters. *E. coli* expressing either the *E. lignolyticus eilA* pump gene **(a)** (see also Chapter 1,), or the *S. typhimurium* homologue *smvA* **(b)** from their native promoters either constitutively (green lines), or under the control of their transcriptional regulators EilR or SmvR (red lines). Gene copy number was directed by using either low copy plasmid (SC101 origin of replication, light green or light red) or high copy number plasmids (colE1 origin of replication, dark green or dark red lines). Cells were grown in medium without [C_2mim]Cl (upper panels) or with 270mM [C_2mim]Cl (lower panels). Curves indicate the mean values of biological duplicate measurements. Note that data represented in (a) and (b) are from two independent experiments, and absolute values cannot be cross-compared. Since the focus of these experiments is on toxicity of the pumps and not on [C2mim]Cl, a negative control strain lacking an efflux pump was not used here.

As previously observed with EilR (Chapter 1, Figure 2), the highest fitness of a cell is established when a substrate-responsive transcriptional regulator keeps expression of the pump at adequate levels. In such a way, the cumulative stress caused by the inhibitor and efflux pump overexpression are kept at a minimum. In the case of the *Salmonella* efflux mechanism, where the pump gene is driven from a relatively strong promoter, the importance of transcriptional regulation becomes even more noticeable.

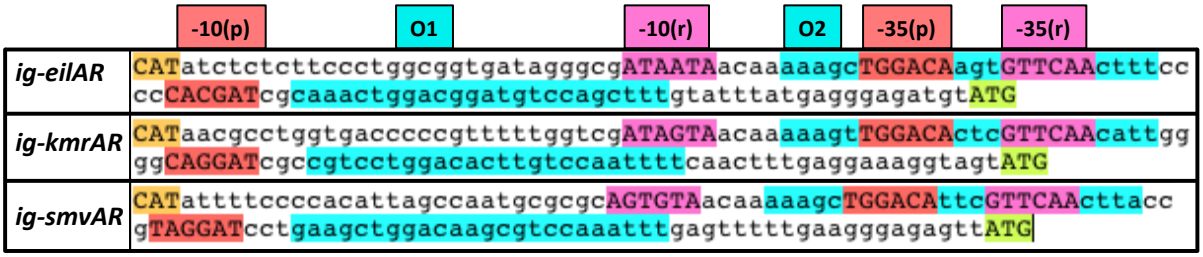


Figure 21 | Intergenic regions with regulatory elements. The 117-119 basepairs intergenic regions of the genes encoding efflux systems in *E. lignolyticus (ig-eilAR*, 117 basepairs*), K. oxytoca (ig-kmrAR*, 119 basepairs*)* and *S. typhimurium (ig-smvAR*, 117 basepairs*)* are shown with their regulatory elements. Colored boxes indicate the -10 and -35 promoter hexamers for the repressor gene ("r", in pink) and for the efflux pump gene ("p" in red) as well as the repressor binding sites O1 and O2 (blue). Start codons are shown for the pump gene (light green) and the divergently transcribed repressor gene (orange).

2.3.4.2 Homologous repressors exhibit a differential sensitivity to their ligands

In this experiment, the *eilR* repressor gene on the biosensor plasmid was replaced with its homologs *smvR* (locus tag: STM1575) from *Salmonella typhimurium* LT2 (Santiviago *et al.*, 2002) and *kmrR* (locus tag: G360DRAFT_2295) from *Klebsiella oxytoca* BRL6-2 (Woo *et al.*, 2014). All three regulators repress the p1 promoter (Figure 22), which indicates that the DNA-binding domains of three repressors are functionally conserved by having a similar affinity to the (truncated) consensus operator in the p1 promoter. Affinity of EilR and KmrR to the previously identified ligands [C_2mim]Cl, crystal violet and acridine orange, is similar. In contrast, SmvR is less sensitive to these compounds, showing only a weak ability to de-repress the p1 promoter.

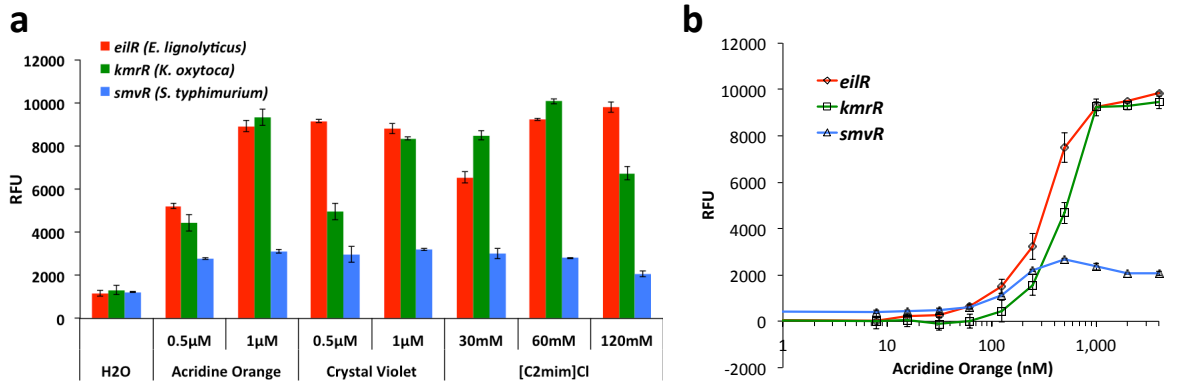


Figure 22 | Response of homologous repressors to EilR ligands. On the reporter plasmid that contains the promoter p1, the *eilR* gene was replaced with its homologues *smvR* (from *Salmonella typhimurium* LT2) or *kmrR* (from *Klebsiella oxytoca* BRL6-2) (see **Figure 18**) **(a)** Fluorescence measurements of *E. coli* harboring the three different reporter plasmids show that the homologous repressor proteins, SmvR and KmrR repress RFP expression, which indicates that they bind to the truncated eilO operator located in the p1 promoter. These proteins have different sensitivities to the EilR ligands acridine orange, crystal violet and [C_2mim]Cl. **(b)** Response of the three homologous repressors to different concentrations of acridine orange. Values and error bars in (a) and (b) represent the means and standard deviation of measurements from two independently grown cultures.

2.4 Conclusions & Outlook

2.4.1 What determines the affinity of a ligand to EilR?

All identified substrates of EilR and EilA are hydrophobic cations, with the positive charge located on the nitrogen atom. While the monovalent cation crystal violet has a high affinity to EilR, its uncharged form, leucocrystal violet, is incapable of inducing EilR-mediated gene expression. This demonstrates that a positive charge is crucial for binding to EilR. Based on my observations made with a range of several triarylmethane cations, I imagine that molecules like crystal violet have the ideal size to clamp themselves into the binding pocket. This enables them to constantly "push the trigger" that causes the conformational change of EilR needed for its release from the operator. Dimethylamine groups, for example, might function as anchor points. With fewer anchors, as with malachite green, the ligand tends to wobble in the binding pocket, resulting in a slacker interaction with the switch residues. At the extreme, larger molecules, such as victoria blue B might be too bulky to fit properly inside the pocket, thereby preventing them form effectively reaching the trigger-switch.

Since many of these statements on EilR-ligand interactions are hypothetical, structural data of EilR bound to these ligands (co-crystallization attempts are ongoing) would help to answer questions, such as: Do structurally dissimilar ligands, for example acridines and triarylmethanes, bind to different sites within the same pocket? Which residues cause the conformational switch? It would also be interesting to test additional triarylmethanes in order to scan the architecture of the binding pocket.

In the meantime, inclusion of structural data from other multidrug-binding repressors that share ligands with EilR might reveal commonalities with respect to a ligand binding mechanism. For example, the structure of the QacR repressor from *Staphylococcus aureus* in complex with the EilR ligand crystal violet has been characterized (Schumacher *et al.*, 2001). In addition, the comparisons made with three homologous repressors could be helpful for characterizing the ligand binding domains and identifying amino acid residues that interact with the cationic ligands. For example, common residues that are not present in the more inert SmvR might establish the similar sensitivity of EilR and KmrR to certain ligands.

2.4.2 Do metabolites act as natural effector molecules?

All currently known ligands of the *E. lignolyticus* EilR repressor and its cognate efflux pump EilA are synthetic substances that are not encountered in tropical rain forest soil, the environment from where the host bacterium was isolated. This behavior aligns with the knowledge of related multidrug efflux mechanisms, for which nearly all known substrates are xenobiotic (Steve Grkovic *et al.*, 2003; Paulsen *et al.*, 1996; Santiviago *et al.*, 2002).

Based on the observations made in the activity-based approach (Figure 16b), it appears that none of the components in LB induce the EilR-mediated promoter. However, it is possible that *E. coli* metabolizes one or more components of this complex medium, generating certain metabolites to which EilR responds, which was displayed by higher RFP expression. Since growth under high pH conditions resulted in an increased promoter activity (Figure 16a), it is reasonable to hypothesize that the putative inducing metabolite is produced at larger quantities at elevated pH. Another factor that determines the activity of the putative inducing metabolite is its charge. As I have shown with leucocrystal violet (Figure 15), the charge of an EilR ligand plays an essential role in EilR-binding. Thus, the putative metabolite might possess a suitable charge for inducing the promoter at elevated pH (e.g. as a monovalent cation), while the metabolite is inert to EilR at lower pH (e.g. as a divalent cation).

The current thinking about metabolite(s) that are only produced under certain growth conditions as potential EilR (and EilA?) ligands are currently of mostly speculative nature. Activity-based assays of fractionated extracts from cells that were grown in complex medium at different pH levels are a promising and unbiased approach to continuing the search for metabolites as natural substrates for this multidrug efflux system. The developed TLC assay is simple and sensitive, in which very small sample quantities are sufficient in order to detect a fluorescent output. Alternatively, extract fractionation by HPLC methods might also prove to be a suitable approach. Subsequent analysis of the active fraction by mass spectrometry, after subtracting the peaks that occur in all fractions, could lead to the detection of inducing metabolites.

2.4.3 Ligand affinity in EilR and EilA is not always harmonized

In the case of $[C_2mim]Cl$, EilR and EilA act as partners, with EilA exporting the inhibitor and the substrate-responsive EilR minimizing stress caused by pump

overexpression (Chapter 1,) (Wagner *et al.*, 2007). However, such a harmonized behavior is not always the case.

Crystal violet was also part of the inhibitor library that was screened in Chapter 1. At that time, this compound did not attract attention, since the constitutively expressed EilA pump did not increase observable tolerance in *E. coli* to inhibitory concentrations (ca. 20μM). Thus, even though nanomolar, subtoxic concentrations of this compound are sufficient to fully de-repress the EilR-regulated *eilA* promoter in the native system, the resulting expression of EilA does not have any beneficial effect on the host but rather creates a burden due to the previously documented stress due to pump overexpression. A contrary situation pertains in the case of methyl viologen, where the repressor poses an obstacle for the efflux of this inhibitor: While methyl viologen is a substrate of EilA, it has only a weak ability to release EilR from its operator. In this case, de-repression of the *eilA* promoter is not sufficient to provide the pump levels needed for the establishment of the tolerance phenotype.

The two contrasting scenarios found with crystal violet and methyl viologen underline the fact that the components of an efflux system, namely the transcriptional regulator and the transporter, do not always cooperate. Since none of the tested xenobiotic compounds are likely to be encountered in the host's natural environment, transcriptional regulation is not always optimized for their efflux.

2.4.4 Importance of transcriptional autoregulation

The results presented in Figure 20 and Figure 22 attempt to show how the two organisms, *E. lignolyticus* and *S. typhimurium,* use two different strategies that both lead to a balanced regulation of two comparable efflux pumps [1]: The weak pump promoter in *E. lignolyticus* requires full de-repression in order to establish the tolerance phenotype, which is provided by a sensitive regulator. In *S. typhimurium* on the other hand, a regulator that is less responsive to the pump substrate is preferable. In this case, partial de-repression of the strong promoter is sufficient to establish tolerance, while excessive SmvA expression can still be avoided.

However, there are three important points that need to be considered in this hypothesis: First of all, the tested ligands that de-repress the p1 promoter are xenobiotic compounds. There might be other (natural) molecules that are more

[1] Douglas Higgins (a Post-Doc in our group) recently expressed *eilA* and *smvA* at equal levels from an independent promoter, and found that these two pumps behave comparably in terms of [C_2mim]Cl transport and intrinsic toxicity.

effective in releasing SmvR from its operator. Secondly, although the three tested repressor proteins have a similar affinity to the consensus operator, there might be operator variants (including the native operators) that enable a better release of SmvR when bound to an effector molecule. Thirdly, constitutive expression of the repressors, which is the case in the described experiments with the biosensor, does not correspond to a natural setting. In the native alignment, the promoter for the *eilR* repressor gene, as well as for *smvR* and *kmrR,* contain an operator (Figure 21), resulting in negative transcriptional autoregulation of the repressor gene. In this way, repressors are expressed at levels just sufficient to occupy the operators for repression in the absence of an inducer molecule (Wu & Rao, 2010). Thus, inducers trigger expression of both the pump and the repressor itself.

Negative autoregulation is a very effective feature used to fine-tune mRNA synthesis. It accelerates response time, reduces variation in expression levels and linearizes the concentration-response (Nevozhay *et al.*, 2009; Wu & Rao, 2010). These are important factors for a strict transcriptional regulation of an efflux pump, which needs to perform its task immediately after a toxic substrate enters the cell. In the efflux systems described here, the genes encoding the pump and its transcriptional regulator are aligned divergently, with each gene being driven by its own promoter (Figure 18 and Figure 21). Such a configuration, which is common among repressors of the TetR-family and their cognate genes (Ahn *et al.*, 2012), results in independent expression levels, minimizing the amount of repressor needed to control pump expression (Wu & Rao, 2010). In this way, the above described negative implications of unequal substrate specificities of the repressor and the pump can be balanced out to a certain degree by autoregulation of the efflux system. In addition, the efflux system can be activated at very low inhibitor concentrations, further increasing sensitivity.

Comparison of both repressor and pump mRNA levels of the three efflux systems under repressed and induced conditions would provide further insight as to if, and to which degree differential sensitivity of the repressors to their ligands affects regulation of their cognate pumps.

2.4.5 Next steps

Being intrigued by the high affinity of EilR to its consensus operator and some of the identified ligands, I will pursue these insights in the next chapter, which deals with the development of inducible gene expression systems.

2.5 Experimental procedures

Purification of the EilR protein

The *eilR* gene was cloned into an expression plasmid (pBR322 origin of replication, kanamycin resistance gene), after an IPTG-inducible "T5"-promoter (pN25). *eilR* was fused to plasmid-borne DNA encoding a C-terminal TEV-protease site linked to the CPD domain and a His-tag (Shen *et al.*, 2009). Detailed, purification procedures, which were optimized for crystallization, are described below. Major differences for the purification of EilR used for EMSA: EilR was expressed as a fusion to the His-tagged CPD-domain, linked by a TEV-protease cleavage site, resulting in high initial protein titers. However, downstream processing after cleaving off the CPD domain and subsequent separation on an ion exchange column resulted in instable EilR protein that was sufficient for bioassays but not for structural studies.

Purification workflow for crystallization [1]: EilR carried an N-terminal His-tag, linked by a TEV protease site. Freshly transformed cells were grown in terrific broth (kanamycin 50mg/mL) to early log phase prior to inducing expression with 500μM IPTG. Cultures were grown at 18°C and 200 rpm in baffled flasks for ca. 60 hours. After removing growth media by centrifugation, the cell paste was lysed in 50mM Tris, pH 8.0, 600mM NaCl, 50mM glutamine, 50mM arginine, 10mM $MgCl_2$, 0.5mM DTT. The EilR lysate was loaded onto the 5mL Histrap column on an ÄKTA FPLC using the sample pump. The column was washed to baseline prior to elution with a gradient of 2-99% buffer 'B' in 20 column volumes. Buffer 'A' was the same as the lysis buffer. Buffer 'B' was prepared by adding 1M imidazole to buffer 'A'. The cleanest elution fractions were pooled and dialyzed back to the lysis buffer with the addition of TEV protease in order to remove the His tag. The cleaved material was put through a 1ml Histrap column on the bench to separate the cleaved EilR from the other components. The purified EilR was then dialyzed against 50mM Tris, pH 8.0, 150mM NaCl, 50mM glutamine, 50mM arginine, 10mM $MgCl_2$, 0.5mM DTT prior to concentration for crystallographic studies.

Electrophoretic mobility shift assays (EMSA)

For the initial EMSA, the full-length *eilAR* intergenic region as well as three truncated versions, were produced by PCR amplification (Table 2) and subsequent column purification. The shorter DNA duplexes used in the subsequent assays were ordered as complementary single strand oligonucleotides, which were then annealed for 10

[1] By Andy DeGiovanni (JBEI, Technology Division)

minutes at 95°C in the presence of 10mM $MgCl_2$. DNA was incubated with or without a 5- to 10-fold molar excess of purified EilR protein at 37°C in Fast-Digest buffer (Life Technologies) for 30 minutes. Following the addition of a 1:10'000-fold dilution of Gel-Red DNA staining dye (Biotium Inc.), samples were run at 100V in a 1% to 1.5% TAE agarose gel. Mobility shifts were detected using a gel imager. The operator sequences were incubated as duplex DNA without or with a 5x molar excess of purified EilR-protein for 30 minutes at 37°C and run for 30 minutes at 90V in an 1.5% agarose gel.

Table 2 | Oligonucleotides used for the generation of full-length and truncated intergenic regions

IG-3'-R	ACATCTCCCTCATAAATACAAAGCTGGACATC
IG117-F	ATCTCTCTTCCCTGGCGGTGATAG
IG92-F	GCGATAATAACAAAAAGCTGGACAAGTGTTC
IG67-F	GTGTTCAACTTTCCCCCACGATC
IG44-F	GCAAACTGGACGGATGTCCAGC

Identification of the consensus operator [1]

In order to identify putative eilO-sites, we first examined the neighborhood of *eilR* and *eilA* homologs in gamma-proteobacteria using pre-computed gene trees available in MicrobesOnline (Dehal *et al.*, 2009). Configurations where homologs of *eilR* form a divergon with homologs of *eilA* were collected and intergenic regions were extracted for further analysis. To improve specificity of motif reconstruction, we filtered out intergenic regions with more than 90% of sequence similarity using Jalview (Waterhouse *et al.*, 2009). The resulting set of non-redundant intergenic regions was used to identify putative TFBS motifs by MEME (Bailey & Elkan, 1994). The MEME algorithm was applied with default parameters, but we restricted the types of motifs to palindromes only, and allowed to search any number of site repetition on the same strand. The motif with the lowest E-value was considered a putative eil-operator.

Construction of the reporter plasmid

In a first step, the *eilR* gene was cloned after a weak constitutive promoter on pFAB5088, a medium copy plasmid (p15A as origin of replication), containing genes encoding kanamycin resistance and a monomeric red fluorescence protein (RFP) as reporter (Mutalik *et al.*, 2013). The *eilR* gene was cloned into the linearized vector backbone by isothermal DNA assembly (Gibson *et al.*, 2009), following the

[1] By Pavel Novichkov (LBNL)

manufacturer's (New England Biolabs) instructions. The resulting plasmid, pFAB_eilR was then used as template to generate a promoter library of randomized -10 and -35 regions. Primers were designed in a way to fit a truncated consensus eil-operator into a 17bp spacer region between the -35 and -10 sites. Since one base pair coincided with the -10 region, the truncated operator contained 18 base pairs. The *eilR*-containing template plasmid was PCR amplified with the primers, eilO-pFAB_random_f and eilO-pFAB_random_r (Table 3), to generate the linearized plasmid with half an operator at each end. This PCR product was then phosphorylated and circularized by self-ligation to create the randomized promoter library. In a next step, these randomized plasmids were transformed into the chemically competent *E. coli* strain, DH10B. [C_2mim]Cl-tolerance of these cells was conferred by the expression of a constitutively expressed *eilA* gene on a low copy plasmid (SC101 origin of replication, chloramphenicol resistant). Transformed cells were plated on 200x200mm Luria-Bertani agar plates supplemented with kanamycin (50mg/L) and chloramphenicol (12.5mg/L) and incubated at 37 °C overnight. One hundred and thirty six colonies were transferred separately into 96-deep-well microtiter plates and grown in EZ-Rich media containing 0.2% glucose and 10μM IPTG either without or with 300mM [C_2mim]Cl. Promoter regions on candidate reporter plasmids were PCR amplified directly from lysed cells and sequenced from a reverse primer annealing in the *rfp* coding region.

Table 3 | Oligonucleotides used for cloning *eilR* into the pFAB5088 and the subsequent generation of a randomized promoter library containing the eil-operator.

Cloning *eilR* into pFAB5088	eilR_f	ccggcaagaaagccttacgaaaataactcaagctgaataacgtgctgc
	eilR_r	ggataaggaggtgacaatatgggctatctgaatcgcgaagaacg
	pFAB5088_f	tcagatagcccatattgtcacctccttatccacacattatacgagcc
	pFAB5088_r	agttattttcgtaaggctttcttgccggaattcgacg
For randomized promoter library	eilO-pFAB_random_f	tgtccaacTANNNTgtgtggagggcccaagttcac*
	eilO-pFAB_random_r	cgtgtccaaNNNCAAgttatgcagcaacgactcatagaaagc*

* eil-operator marked in red, promoter boxes are in upper case

RFP fluorescence measurements

End-point measurements

Samples (150-200uL) of cells were transferred into 96-well black microtiter plates with clear bottom and absorbance was measured at 600nm using a Spectramax microplate reader (Molecular Devices). For RFP fluorescence, excitation was at 585nm and emission at 607nm. Alternatively, endpoint measurements were

performed in an Infinite F200pro microplate reader (Tecan), using the same gain and excitation/emission settings as for the kinetic measurements (see below).

Flow cytometry

Stationary phase cultures were diluted 1:1000 in PBS buffer. Cells were counted (5000 events per sample) using a Guava easyCyte (Millipore) flow cytometer by forward and side scatter acquisition, and the cellular accumulation of RFP was measured by fluorescence intensity. Data acquisition was performed using InCyte software version 2.2 (Millipore). A strain of *E. coli* lacking RFP was used to subtract background fluorescence.

Kinetic assays

Kinetic assays were performed in an Infinite F200pro microplate reader (Tecan). Cells were grown in black, clear bottom 96-well microtiter plates at 200μL/well. Cell density was measured at 600nm. RFP was excited at 575nm (20nm bandwidth) to measure emission at 620nm (20nm bandwidth). Gain was typically adjusted manually to a value of 40.

Imaging of RFP fluorescence

Cells were recorded in a gel imaging station (HP Alphaimager, Protein Simple) under the Cy3 filter (537nm). Exposure time and aperture were adjusted according to RFP intensity.

Screening for EilR-ligands

All chemicals were purchased from Sigma-Aldrich. Stock solutions were prepared in water at a concentration of 2mM (chemical dyes) or 4M (ILs, [C_1mim]Cl, [C_2mim]Cl, [C_4mim]Cl). Chemically competent *E. coli* DH10B were transformed with the reporter plasmid pTR_p01 and grown on agar plates containing Luria-Bertani medium supplemented with kanamycin at 50mg/L. A colony was picked and grown to stationary phase as seed culture.

Unless stated otherwise, aqueous ligand solutions were prepared as a 2-fold dilution series in 96-deep well microtiter plates. A 50-fold dilution of seed culture was suspended in a 2-fold concentrated EZ-Rich defined medium (kanamycin, glucose) and added to the ligand solution to a final volume of 600μL. Cells were grown in a rotatory lab device at 900rpm at 37°C for 18hours.

Preparation of cell-extracts

E. coli DH10B cultures were grown in 40mL of either EZ-Rich medium containing 0.2% glucose or in LB broth for 19h at 37°C. Cultures were then transferred to a 50mL falcon tube and centrifuged for 10 minutes at 3000rpm at 4°C. After removal of the supernatant, the cell pellet was resuspended in 40mL ice-cold PBS buffer and washed by short agitation. The buffer was removed after centrifugation for 10 minutes at 3000rpm at 4°C. The cell pellet was resuspended in 7mL ice-cold methanol by pipetting up and down and by vigorous vortexing for 1 minute. 7mL ice-cold water was added and the samples was again vortexed for 1 minute. To enhance lysis, the sample was then flash-frozen in liquid N_2. After thawing, the extract was distributed to 2mL eppendorf tubes and centrifuged at maximum speed at 2°C for 5 minutes. The supernatant was transferred to a protein purification spin-column (Sartorius Vivaspin 20) with a cut-off size of 3kDa and centrifuged at maximum speed at 2°C. Methanol in the small molecular weight flow-through was removed in a speed-vacuum centrifuge at 45°C. Samples were then flash-frozen in liquid N_2 and lyophilized to remove the remaining water. The oily extracts were stored at 4°C.

Thin layer chromatography (TLC)

C18-silica plates were used as a stationary phase. 4-fold dilutions of cell extracts or 2-fold concentrate of LB-media were applied to TLC plates at 1μL/dot. A mix of methanol/ethyl acetate (70:30) was used as a mobile phase. After fractionation of the extracts, plates were dried and imaged under the Cy3 filter. TLC plates were then submerged in EZ Rich/0.2% glucose medium containing a 1:100 dilution of a stationary phase *E.coli* biosensor seed culture and incubated overnight at 37°C (The seed culture was also grown in EZ Rich/0.2% glucose medium). TLC plates were then again imaged under the Cy3 filter.

Generation of response curves

Concentration-response curves and the corresponding parameters such as Hill-coefficient, EC_{50}, y_{min} and y_{max}, were calculated using the four parameters nonlinear regression function provided by Prism statistical analysis software (Graphpad Inc).

Chapter 3

EilR-regulated expression systems

Image: *E. coli* expressing RFP from EilR-mediated phage promoters after induction with crystal violet

3.1 Summary

Inspired by the results obtained in Chapter 2, I decided to apply EilR-based transcriptional regulation for the development of a gene expression platform. The high affinities of EilR to its operator DNA and to some of its identified ligands provide a good basis for tight transcriptional control. I engineered a range of bacterial promoters by using three different approaches, namely 1) by refashioning the p1 "biosensor" promoter described in section 2.3.2 , 2) by introducing eil-operators into several bacteriophage promoters and 3) by modifying the core promoter sites of the native regulatory DNA region. This generated a set of promoters with very low background activity that are, upon addition of nanomolar concentrations of cationic dyes, inducible over a range of up to 4-5 orders of magnitude, reaching levels comparable to those of the strongest bacterial expression systems. Transferring this expression system from *E. coli* to the soil bacterium *Pseudomonas putida* and the nitrogen-fixing plant symbiont *Sinorhizobium meliloti* illustrates that the developed EilR-mediated promoters are also functional in distantly related bacteria. In the final section, I introduce eil-operators into a yeast promoter to enable EilR-mediated gene expression in an eukaryotic organism. These findings can provide a novel way for orthogonal and tightly controllable high-level gene expression at negligible costs.

3.2 Introduction

Transcription factors and their cognate DNA binding sites can be decoupled from their native context and used to control expression of desired target genes. Thus, these regulatory elements are essential tools for the fundamental understanding of gene function and for regulating expression levels. In biotechnological applications, inducible promoters are essential tools for adjusting expression levels and for decoupling growth from production, which enables the establishment of a healthy culture prior to experimentally directed expression.

Probably the most straightforward way to have full control over the moment and level of induction is to use an inducible promoter that can be exogenously controlled by the addition of an inducer molecule. A promoter is ideally strictly regulatable, showing minimal activity in the repressed state and high transcription rates when induced. At the same time, it should be orthogonal, meaning that it operates in an isolated manner, independent from growth state and media composition and without interfering with the host metabolism or other expression systems. To fulfill these requirements, interaction of the transcription factor with its inducing ligand and its DNA binding site must be very specific. The inducing molecule should (1) not be

toxic at operational concentrations, (2) not be metabolized by the host organism, (3) not affect host physiology but (4) uniquely and exclusively trigger a conformational change of the transcription factor to alter DNA binding affinity.

Prokaryotic genes or operons are often regulated by a single transcription factor that binds to a unique operator in the promoter region, some of which are described on page 9. Such an orthogonal functionality provides a good basis for engineering an inducible expression system that can be controlled by specific effector molecules.

Compared to bacteria, external control over gene expression in eukaryotes is more challenging, since their promoters typically contain multiple binding sites for various transcription factors that activate or repress promoter activity. At the same time, a transcription factor typically influences regulation of more than just one gene. This complex network enables the organism to fine-tune gene expression to its metabolic needs and to external conditions, such as nutrient availability. Due to this complexity, tools to regulate expression levels via orthogonal promoters that do not interfere with the host metabolism are limited in yeast. A prominent example of a complex induction system, the GAL1 promoter, as well as more independent, heterologous alternatives, are described on page 11.

Here, I describe the combination of EilR-mediated control using several less common phage promoters in order to develop an inducible gene expression system in bacteria that is tightly repressible and highly inducible with cationic dyes. I then describe how I inserted EilR-based transcriptional regulation into *S. cerevisiae*. The findings open up a way to use this bacterial system to orthogonally control gene expression in eukaryotic organisms.

3.3 Results & Discussion

3.3.1 EilR-mediated bacterial promoters

This section describes the development and properties of novel EilR-regulated promoters in bacteria.

3.3.1.1 Approach

The promoters were engineered by pursuing three approaches, which are summarized in Figure 23.

Approach 1: Reducing leakiness of the biosensor promoters

Here, the previously described EilR-regulated promoters p1 and p2 (Figure 10) were taken as template to add an additional operator with the intention to reduce basal transcriptional activity. The consensus operator was inserted 6 base pairs downstream of the -10 promoter hexamer in a way that the 5' end of the operator coincided with the characterized transcriptional start site (Mutalik *et al.*, 2013).

Approach 2: Integration of eil-operators into early bacteriophage promoters

To enable rapid gene expression upon infection, bacteriophages efficiently redirect the host transcription machinery towards their genome. Therefore, phage promoters driving the expression of early genes need to be strong and readily recognized by the host RNA-polymerase.

I chose four early promoters from the three coliphages, namely the pA1 promoter from phage T7, the pL promoter from the lambda-phage and the pDE20 and pH207 promoters from phage T5 (Deuschle *et al.*, 1986). With the exception of pL, these promoters have not gained much attention since their characterization. They should not be confused with the commonly used phage promoters termed "T5" and "T7"-promoter (see page 11). In a first step, I inserted the truncated eil-operator site into the 17 base pair spacer region located between the promoter hexamers. With the assumption for tighter repression, promoters were designed in a way that the operator overlapped with the -35 and/or the -10 hexamer. Analogously to approach 1, an additional, full-length operator was then inserted downstream of the core promoter region. In all designs, the 5' end of the operator coincided with the transcriptional start site, which is formed by an adenine in all promoters used (Deuschle *et al.*, 1986).

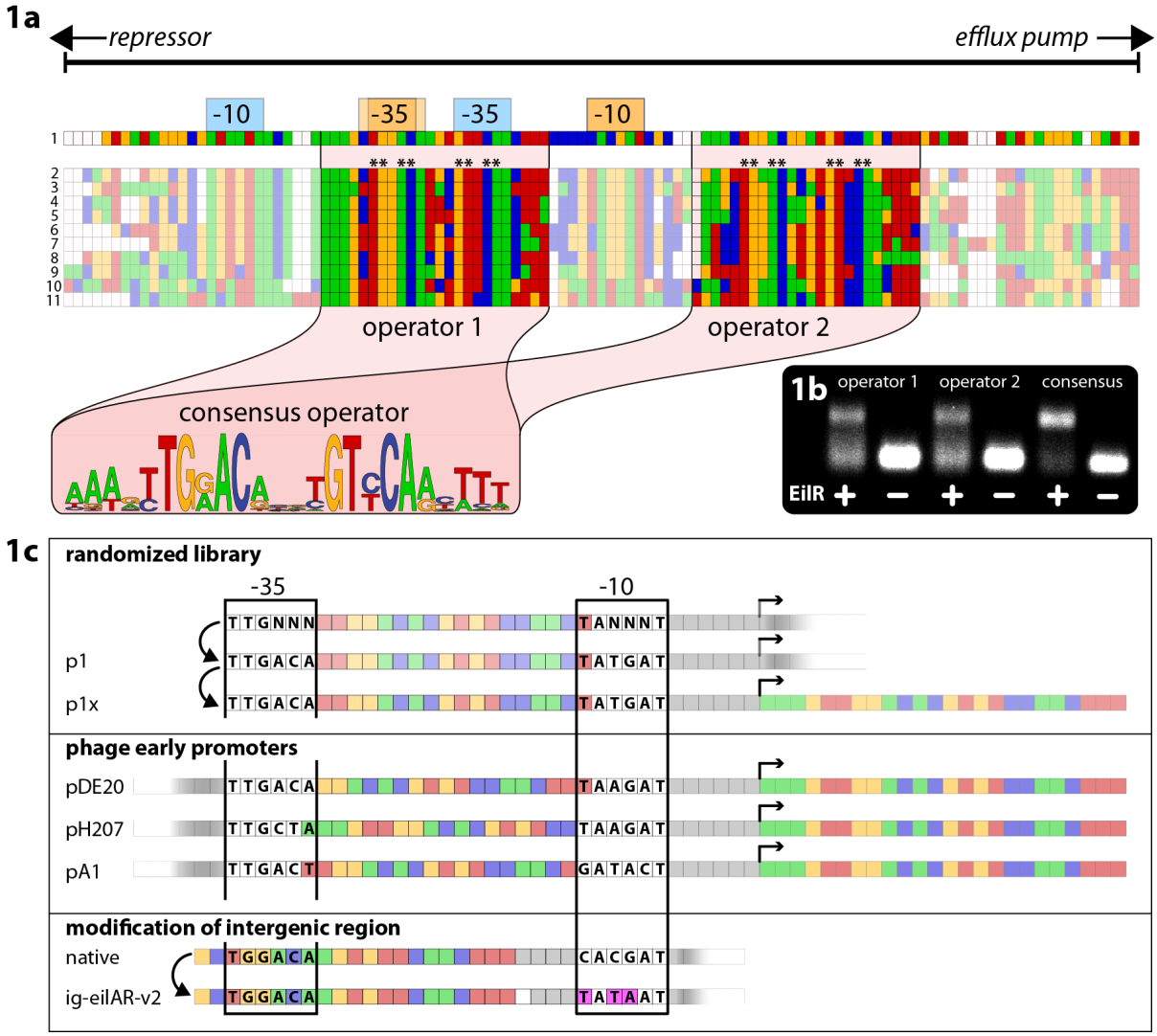


Figure 23 | Overview of the workflow taken to engineer EilR-regulated promoters. **(a)** Alignment of intergenic regions that separate the genes encoding homologous efflux pumps and their cognate repressors in eleven selected bacteria: *Enterobacter lignolyticus* (1); *Citrobacter koseri* (2); *Citrobacter rodentium* (3); *Salmonella enterica paratyphi* (4); *Salmonella enterica arizonae* (5); *Klebsiella pneumoniae* 342 (6); *Klebsiella pneumoniae* NTUH-K2044 (7); *Enterobacter sp.* 638 (8); *Pantoea ananatis* LMG 20103 (9); *Acinetobacter sp.* ADP1 (10); *Acinetobacter baumannii* (11). All intergenic regions contain two repressor binding sites (operators 1 and 2), which enabled the generation of a perfect palindromic 24 base pair consensus operator motif. Base pairs that are conserved throughout the depicted operators are indicated with an asterisk. Predicted promoter hexamers are indicated for the repressor genes (light blue) and for the efflux pumps (orange). Each nucleotide is represented by a colored box: A (green); T (red); G (yellow); C (blue). **(b)** Electrophoretic mobility shift assay indicates that the purified EilR repressor protein has a higher affinity for the consensus operator than for the native *E. lignolyticus* operators (see also p. 39). **(c)** EilR-regulated promoters were developed in three ways: The previously described randomized approach generated promoter p1, into which an additional eil-operator was placed to yield the p1x version. Secondly, a truncated and a full-length consensus operator was inserted into phage-promoters. Thirdly, the native *E. lignolyticus eilA*-promoter, which is located in the intergenic region, was modified, with the mutated nucleotides shown in pink. Arrows indicate the transcriptional start site. Nucleotides belonging to the eil-operator are color-coded. Promoter -35 and -10 hexamers are boxed.

Approach 3: Modification of the native E. lignolyticus eilA promoter

The intention of this approach was to increase strength of the wild-type *eilA* promoter, while keeping the operators intact for repression. The -35 region of the *eilA*-promoter is embedded in the operator sequence. Modification of this site bears the risk that it would affect binding affinity of EilR to its operator, resulting in weaker repression. In contrast, repression should not be affected when the -10 site of the *eilA*-promoter, located in between the two operators, is mutated.

Especially the initial thymine is very conserved in *E. coli* promoters (Lisser & Margalit, 1993), and its presence is important for binding of the RNA polymerase in complex with the σ^{70} "housekeeping" sigma factor (Shimada et al., 2014), but is missing in the wild-type *eilA* promoter. Thus, exchanging three nucleotides of the wild type -10 site (CACGAT) to the consensus sequence of *E. coli* promoters (TATAAT) should result in facilitated transcription. (I later realized that the spacer was accidentally shortened to 16bp instead of maintaining the consensus 17bp, suspecting that this negatively affected promoter strength.)

The promoters were developed and tested in a plasmid-based approach. To avoid potential effects on expression unrelated to the promoter, the plasmid backbones were identical, allowing for cross-comparison.

3.3.1.2 EilR-regulated promoters are tight and strong

Targeted mutations in the native *eilA* promoter towards the *E. coli* consensus sequence resulted in a strong increase of promoter activity (**Figure 24**). However, the highest dynamic range was observed in the developed phage promoters.

Insertion of a single truncated operator into phage promoters enabled only very weak repression. In the presence of crystal violet, these promoters were induced by a factor ranging from 2 to 30. Addition of a second operator downstream of the core promoter region drastically increased repression (Figure 25, Figure 26, Figure 28). Providing an additional EilR binding site at this location efficiently blocked mRNA synthesis. Addition of a second operator also suppressed the leakiness of the p1 biosensor promoter observed under certain growth conditions (see Chapter 2, section 2.3.3.3), resulting in the tightly repressible p1x promoter (Figure 27). Presence of the second operator also raised promoter activity in the de-repressed state. A possible reason for this is further discussed on the next page.

The resulting promoters had a dynamic range of more than four orders of magnitude (Figure 25). The highest dynamic range was observed when the operators were introduced into the pDE20 promoter from phage T5. In the induced state, RFP expression driven by this promoter was higher than the one from the phage RNA polymerase dependent T7-promoter (Figure 29, Figure 30b), which is considered to be the strongest expression system in *E. coli* (Balzer *et al.*, 2013a; Terpe, 2006). Concentrations of crystal violet needed for full induction were dependent on plasmid copy number (Figure 26), which is explainable by the different levels of EilR that need to be saturated by the inducer molecule. In any case, inducing concentrations of crystal violet did not measurably inhibit growth in *E. coli,* as was the case for the inducer acridine orange (Figure 34).

When repressed, basal activity of several eil-promoters was barely detectable at all tested growth stages and media types (Figure 26, Figure 28). In fact, repression was more effective than the one observed with all other tested promoters, including the arabinose-regulated pBAD promoter that is considered to be the tightest existing promoter in *E. coli* (Balzer et al., 2013; Terpe, 2006). Testing the p1x promoter and its inducer acridine orange for orthogonality did not reveal any crosstalk with the three commonly used inducible expression systems (Figure 31).

Increased mRNA stability due to an operator stem-loop?

Insertion of a second consensus operator not only increased repression, but also resulted in elevated RFP levels in all but one of the tested promoters in their induced state. I hypothesize that the higher reporter protein levels were not caused by increased promoter strength but rather by enhanced mRNA stability. In fact, the full-length operator coincides with the transcriptional start site of typical *E. coli* (and most other bacterial) promoters, therefore forming the 5-prime end of the mRNA. The inverted repeats in this 24 base pair perfect palindromic operator have the potential to form a strong RNA stem-loop. Such 5-prime secondary structures were shown to protect mRNA from its degradation by RNases, resulting in higher template levels available for translation (Bouvet & Belasco, 1992; Emory *et al.*, 1992). On the other hand and in contradiction to this hypothesis are observations made with promoter p2, which differs from the p1 version only by three base pairs in the -10 region (Figure 10, p. 43). Here, insertion of an additional operator decreased RFP expression from the fully induced promoter, although the mRNAs transcribed from p1x and p2x are identical. Measuring mRNA levels by reverse transcription PCR after rifampicin-mediated inactivation of RNA-polymerase (Campbell *et al.*, 2001) could be a way to gain insight whether the predicted operator stem-loop increased mRNA abundance by increasing its stability.

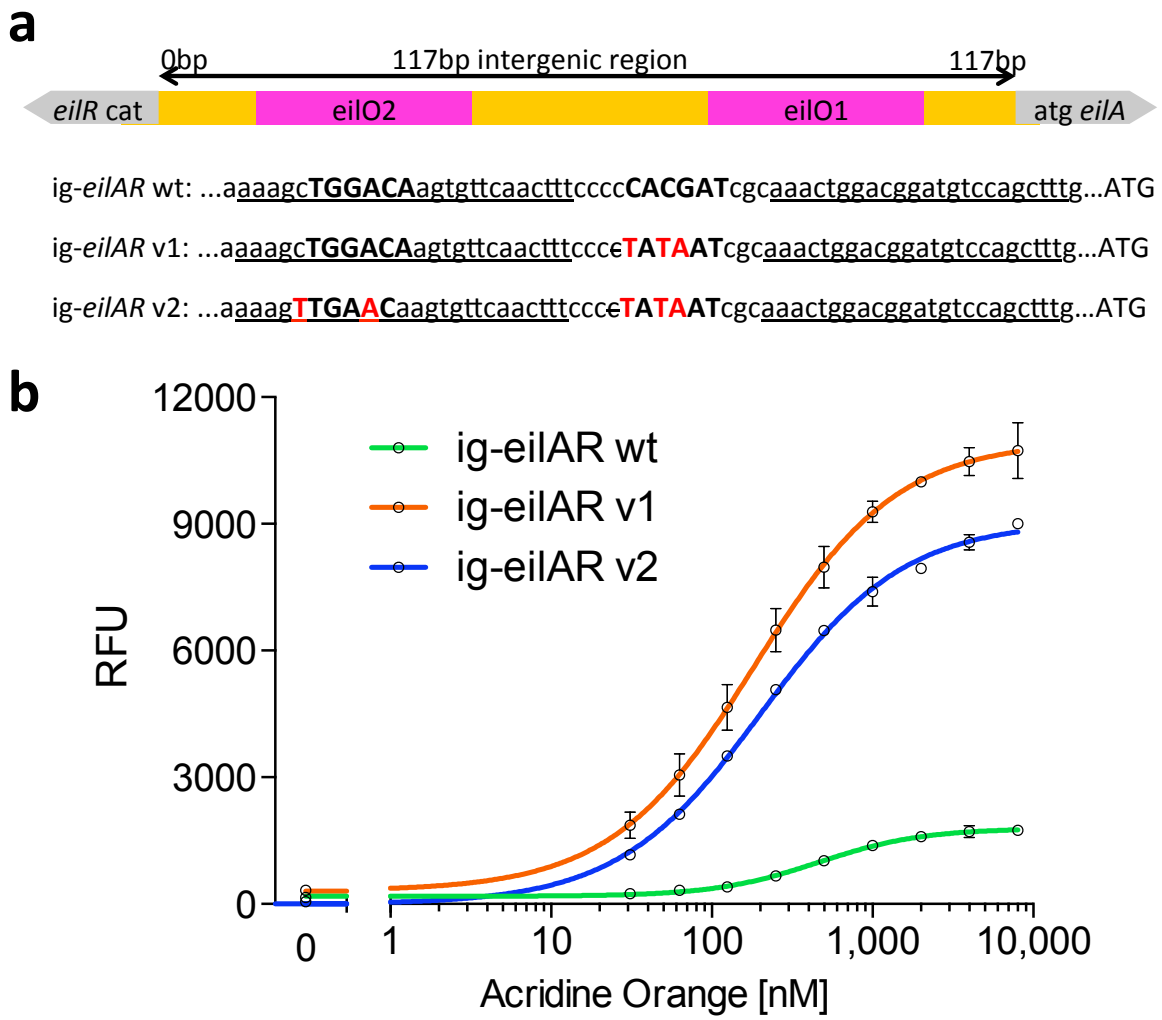
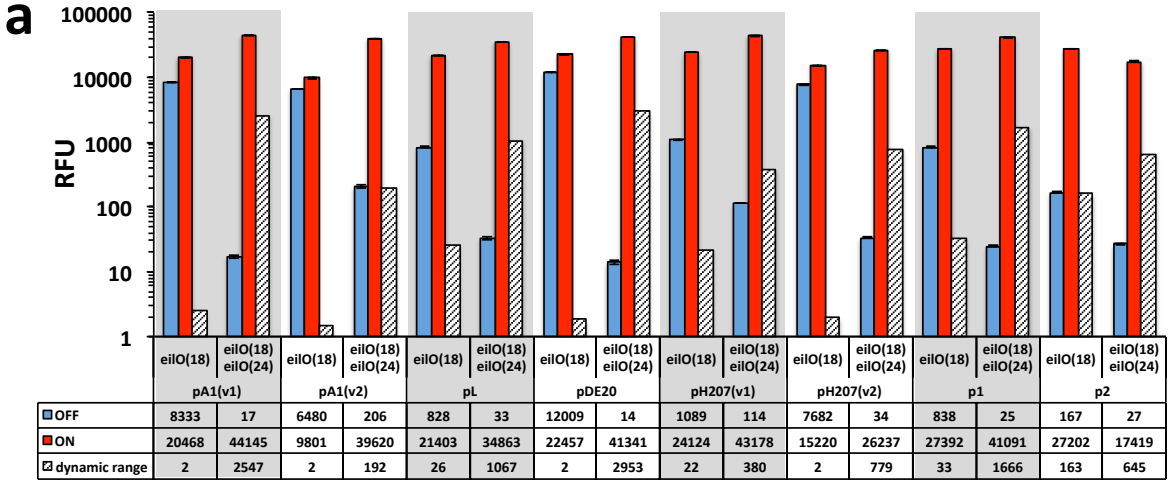


Figure 24 | Modifications in the core promoter regions of the native E. lignolyticus eilAR intergenic sequence increases promoter strength. (a) The eilA-efflux pump gene was replaced with rfp while the eilR repressor gene and the upstream intergenic region containing two operators (underlined) and promoter regions (upper case) were maintained. Base pairs labeled in red in the promoters ig-eilAR v1 and v2 indicate modifications in the -35 and the -10 regions of the E. lignolyticus wild type eilA promoter (ig-eilAR wt). **(b)** Fluorescence measurements of stationary phase E. coli cultures show that the modifications in versions v1 and v2 result in increased RFP expression compared to the wild type. Values and error bars represent the means and standard deviation of measurements from two independently grown cultures.

a



	pA1(v1)		pA1(v2)		pL		pDE20		pH207(v1)		pH207(v2)		p1		p2	
	eilO(18)	eilO(18) eilO(24)	eilO(18)	eilO(18) eilO(24)	eilO(18)	eilO(18) eilO(24)	eilO(18)	eilO(18) eilO(24)	eilO(18)	eilO(18) eilO(24)	eilO(18)	eilO(18) eilO(24)	eilO(18)	eilO(18) eilO(24)	eilO(18)	eilO(18) eilO(24)
OFF	8333	17	6480	206	828	33	12009	14	1089	114	7682	34	838	25	167	27
ON	20468	44145	9801	39620	21403	34863	22457	41341	24124	43178	15220	26237	27392	41091	27202	17419
dynamic range	2	2547	2	192	26	1067	2	2953	22	380	2	779	33	1666	163	645

b

origin	promoter name	promoter activity (RFP) uninduced	promoter activity (RFP) induced	dynamic range
T7 phage	pA1	0.29 ±0.1	3105 ±337	10748
T5 phage	pH207	0.2 ±0.06	2971 ±231	14869
T5 phage	pDE20	0.13 ±0.05	3047 ±124	23560
Lambda phage	pL	1.14 ±0.13	1263 ±32	1106
synthetic	p1x	0.48 ±0.1	2464 ±101	5113

Figure 25 | Properties of EilR-regulated promoters originating from three different coliphages or synthesized de novo by a randomized library approach. (a) Normalized fluorescence of *E. coli* cultures expressing RFP from EilR-regulated promoters containing one ($eilO_{18}$) or two operators ($eilO_{18}$&$eilO_{24}$)(see **Figure 23** & 24). (b) Mean single cell fluorescence of *E. coli* expressing RFP from five selected EilR-regulated promoters that all contain two operators. Cells were grown to stationary phase in EZ Rich/0.2% glucose (a) or in TB (b) and expressed RFP from medium copy (p15A) plasmids. Values and error bars represent the means and standard deviation of measurements from three independently grown cultures.

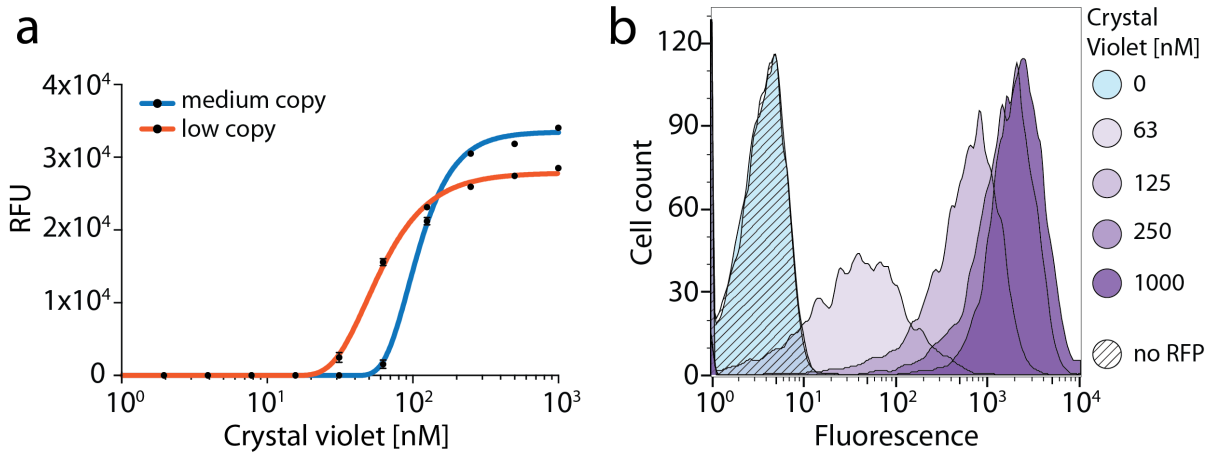


Figure 26 | Dynamic range of DE20-eilOx promoter. (a) Mean RFP-fluorescence emitted by *E. coli* containing either a low copy (SC101, red) or medium copy (p15A, blue) DE20-eilOx reporter plasmid after growth in increasing concentrations of the inducer crystal violet. **(b)** Histogram representing the distribution of single cell fluorescence of *E. coli* driving RFP expression from DE20-eilOx on a medium copy plasmid. Cultures were grown without (blue) or in the presence of increasing crystal violet concentrations (purple). The hatched histogram represents background fluorescence from a control strain lacking the *rfp* gene. Crystal violet was added to EZ-Rich media containing 0.2% glucose at the time of inoculation. Cultures were grown to stationary phase for measurements. In (a), values and error bars represent the means and standard deviation of measurements from three independent experiments.

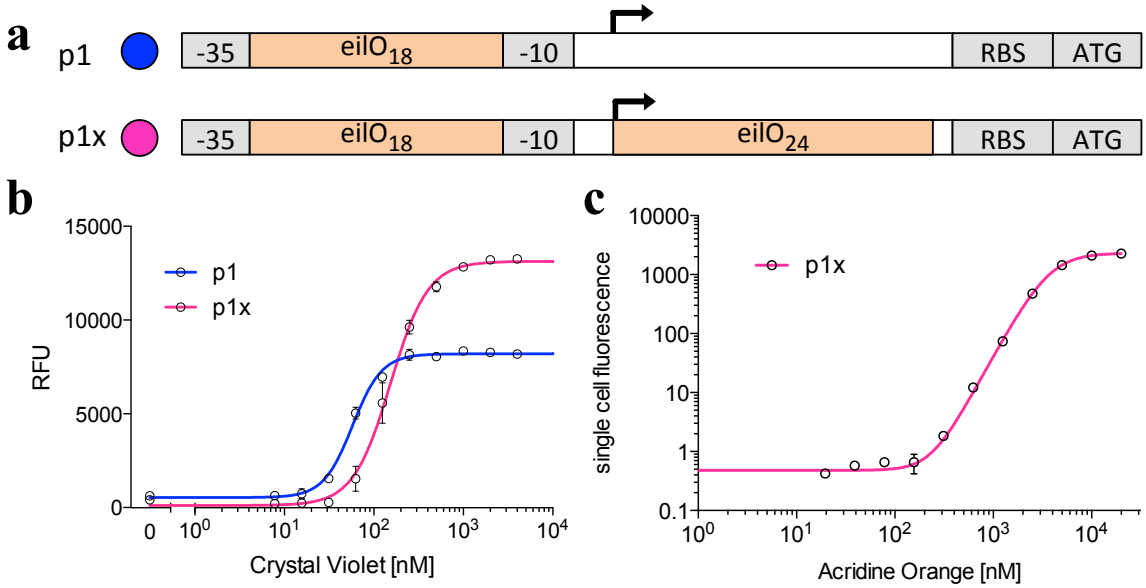


Figure 27 | Addition of a full-length operator increases the dynamic range. (a) The p1 promoter containing a truncated 18bp operator in the spacer region between the -10 and -35 sites (blue) was used to develop the p1x promoter (red) with an additional full length 24bp operator at the transcriptional start (black arrow). **(b)** Single cell fluorescence measurements of *E. coli* expressing RFP from the two promoter versions at different concentrations of crystal violet, indicating that insertion of the full-length operator reduces basal expression in the repressed state. Note that RFP levels were higher in the fully induced p1x. Measurements were taken after cells were grown to stationary phase in EZ-Rich defined medium containing 0.2% glucose. **(c)** Mean single cell fluorescence of *E. coli* expressing RFP from the p1x promoter at different concentrations of the inducer acridine orange. Note the logarithmic scale. Values and error bars represent the means and standard deviation of measurements from two independently grown cultures.

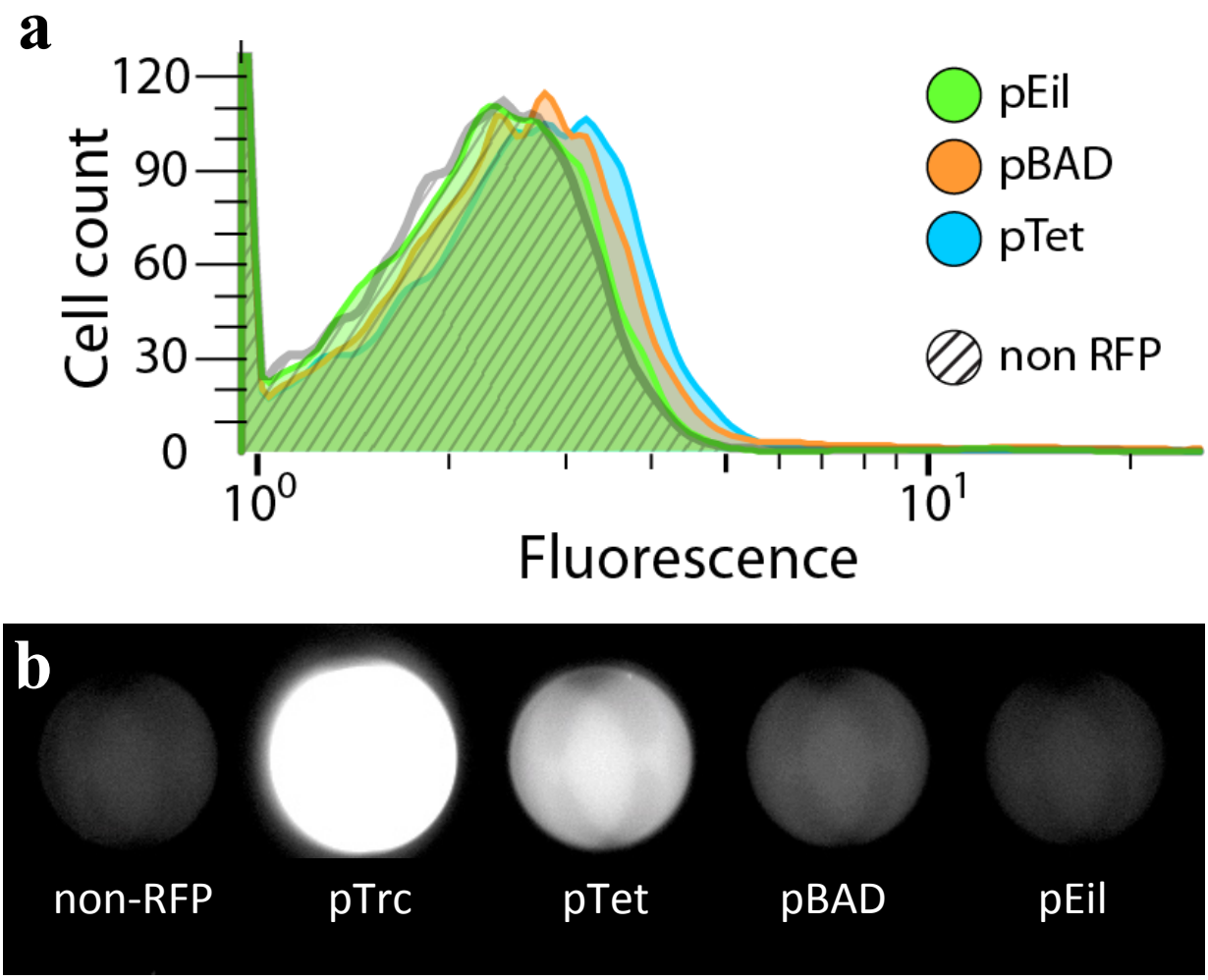


Figure 28 | RFP fluorescence in uninduced *E. coli* resulting from basal activity of different promoters. (a) Distribution of single cell fluorescence of *E. coli* harboring medium copy *rfp*-plasmids (p15A). The hatched histogram represents background fluorescence from a control strain lacking the *rfp* gene. **(b)** Fluorescence of *E. coli* (OD_{600nm}=1) that carry plasmids that all contain the same RBS and origin of replication (p15A), and drive expression of the *rfp*-gene from either pEil (DE20-eilOx in a, p1x in b) or the common pTet (pBBA2k), pBAD (pBBA8k), pTrc (pBBA1k) (T. S. Lee *et al.*, 2011). Fluorescence was imaged under the Cy3 filter. Cells (strain DH10B) were grown in the absence of inducer molecules to stationary phase in either EZ-Rich defined medium containing 0.2% glucose (a) or in terrific broth (b).

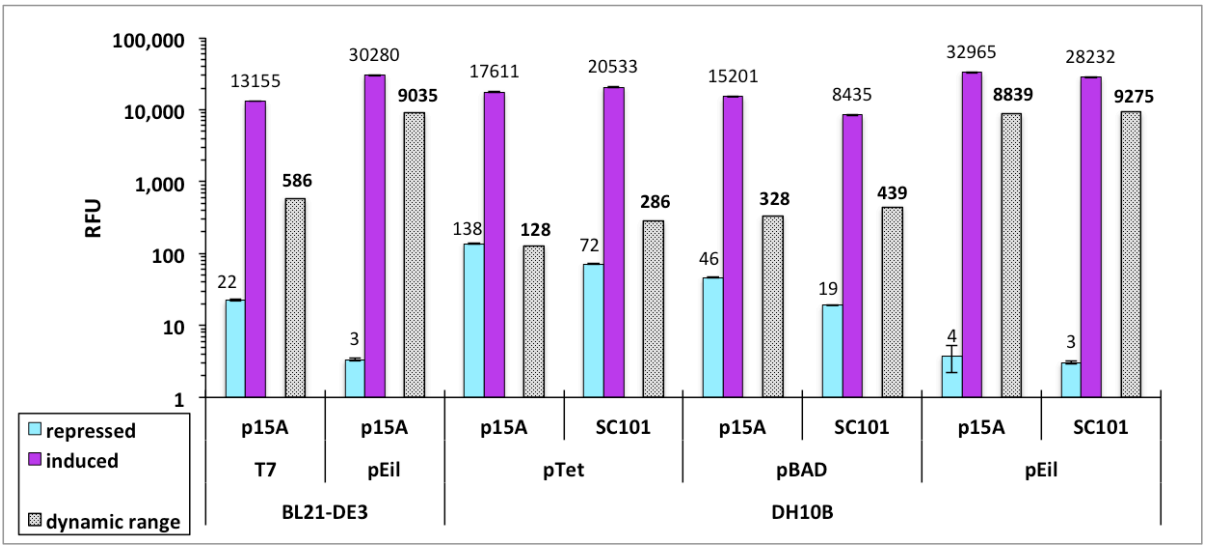


Figure 29 | RFP expression from different inducible promoters. Normalized RFP-fluorescence of two *E. coli* strains containing either low copy (SC101) or medium copy (p15A) *rfp*-plasmids was measured after growth to stationary phase in EZ-Rich defined medium containing 0.2% glucose. Plasmid backbones, including the ribosomal binding site, are identical among the tested variants. T7: pBBA7k (p15A); pTet: pBBA2k (p15A) and pBBS2k (SC101); pBAD: pBBA8k (p15A) and pBBS8k (SC101) (T. S. Lee *et al.*, 2011). Inducers were added at the time of inoculation at the following concentrations: Crystal violet 1µM; IPTG 500µM; anhydrotetracycline 200nM; D-arabinose 10mM. *E. coli* BL21-DE3 were freshly transformed by electroporation prior to inoculation, DH10B were grown to a seed culture from a glycerol stock. Fluorescence values were normalized to cell density and background fluorescence from a non-RFP strain was subtracted. Values and error bars represent the means and standard deviation of measurements from three independently grown cultures.

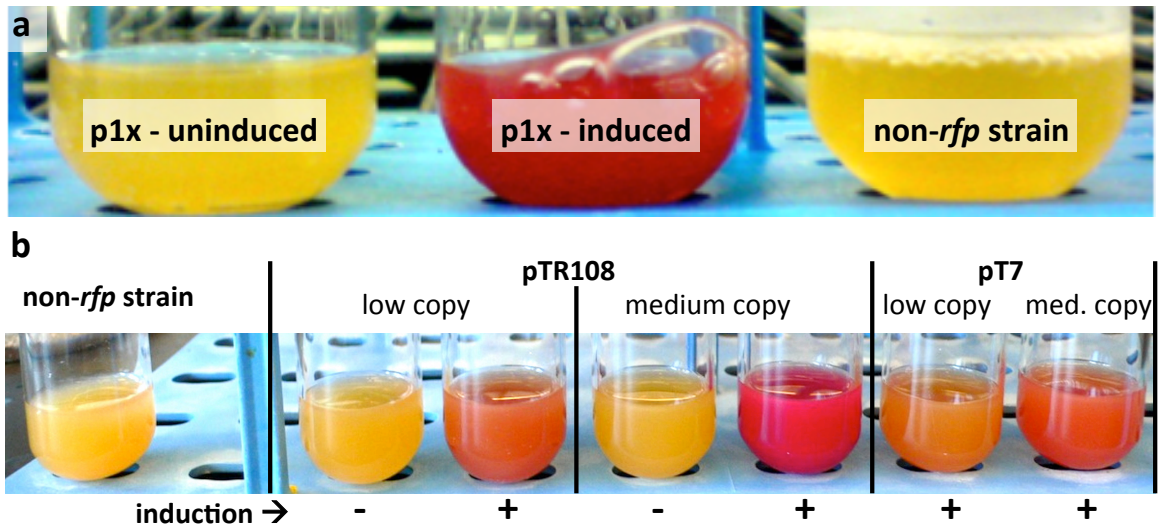


Figure 30 | RFP expression from EilR-regulated promoters. (a) Image of *E. coli* DH10B expressing RFP from the p1x promoter on a medium copy number plasmid. **(b)** Image of *E. coli* BL21-DE3 expressing RFP from either the DE20-eilOx promoter or the conventional, LacI-regulated T7-promoter on low- or medium copy number plasmids. Cultures were induced at early log-phase with either crystal violet at concentrations of 2µM (a) or 500nM (b) or 500µM IPTG (T7), and grown to stationary phase at 37°C for 50h (a) or at 30°C for 20h (b).

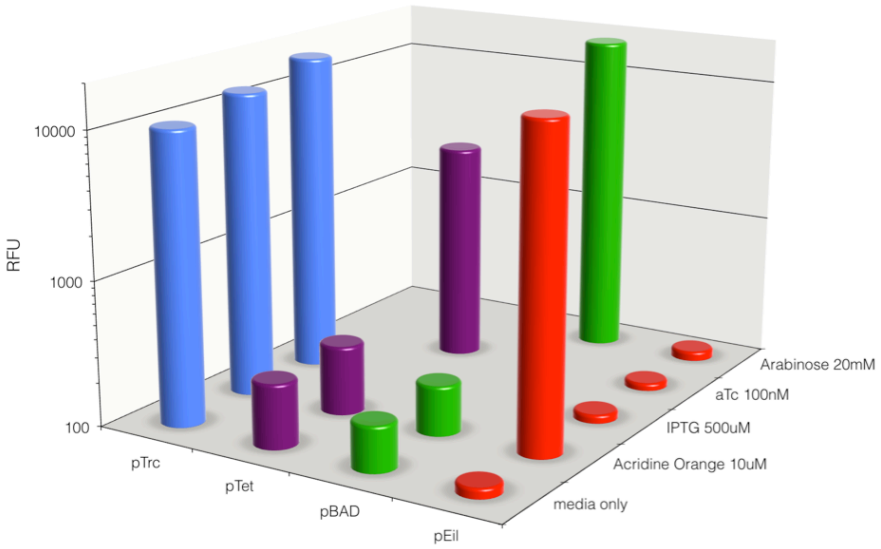


Figure 31 | Orthogonality of the p1x Eil-promoter and its inducer acridine orange. Normalized fluorescence of *E. coli* cells expressing RFP from common inducible promoters and from the EilR regulated p1x promoter (labeled pEil). Cultures were induced in early log-phase of growth (OD=1.3) and grown to stationary phase in terrific broth for 21 hours. *E. coli* background fluorescence was not deducted.

Negative autoregulation versus constitutive repressor levels

Of the developed promoters, the versions that are regulated by a constitutively expressed EilR repressor (phage and modified biosensor promoters) had the lowest basal activity and largest dynamic range.

The constant repressor levels provided by its constitutive expression resulted in a relatively steep concentration-response resembling an ON/OFF switch when the promoters were induced by crystal violet. In contrast, a constellation where the EilR repressed its own transcription, as it is found in the native alignment and its derivatives (ig-eilAR versions), provided a more sensitive, gradual concentration-response (**Figure 32**).

The more sensitive response found in autoregulated promoters could be useful for finely titratable expression. However, the current ig-eiAR versions are very leaky, i.e. they possess a relatively high basal activity in the uninduced state. It would therefore be interesting to combine the favorable characteristics of both alignments, namely to develop an autoregulated, tight and strong promoter version with a high dynamic range. This could be done by replacing the constitutive *eilR*-promoter with a partial intergenic region that contains the eilO2 operator and the native *eilR*-promoter (for annotation of the intergenic region, see Chapter 2, Figure 23). Certain less conserved sequences in the operator and the -10 site can then be randomized to modulate promoter strength and EilR affinity to the operator. For example, increasing promoter

strength of the *eilR* gene should increase repression by providing more EilR protein for repression. The randomized library can then be examined in *E. coli* by FACS, allowing the selection of tightly repressed candidates that do not emit fluorescence.

The choice of inducer molecule is another way to modulate the concentration-dependency of EilR-mediated promoters. The response of the DE20-eilOx promoter to crystal violet was very steep, indicating that this compound triggers a nearly "digital" ON/OFF response. On the other hand, regulation of the same promoter by acridine orange is more gradual, enabling a better handling when intermediate expression levels are preferred (Figure 13, Figure 33).

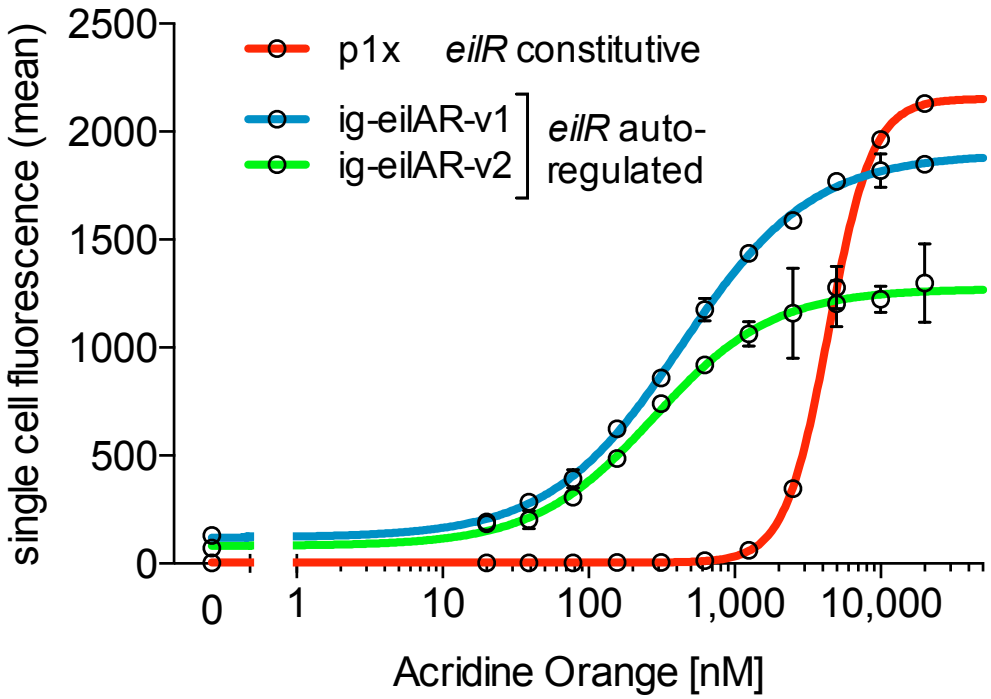


Figure 32 | Repressor autoregulation flattens the concentration-response. Mean single cell fluorescence of E. coli expressing RFP from three different EilR-regulated promoters in the presence of increasing concentrations of the inducer acridine orange. In the p1x version, EilR is constitutively expressed, while in the two modified native regulatory regions (ig-eilAR v1/v2), EilR regulates its own transcription in addition to *rfp*. Values and error bars represent the means and standard deviation of measurements from three independently grown cultures, which were grown to stationary phase in terrific broth.

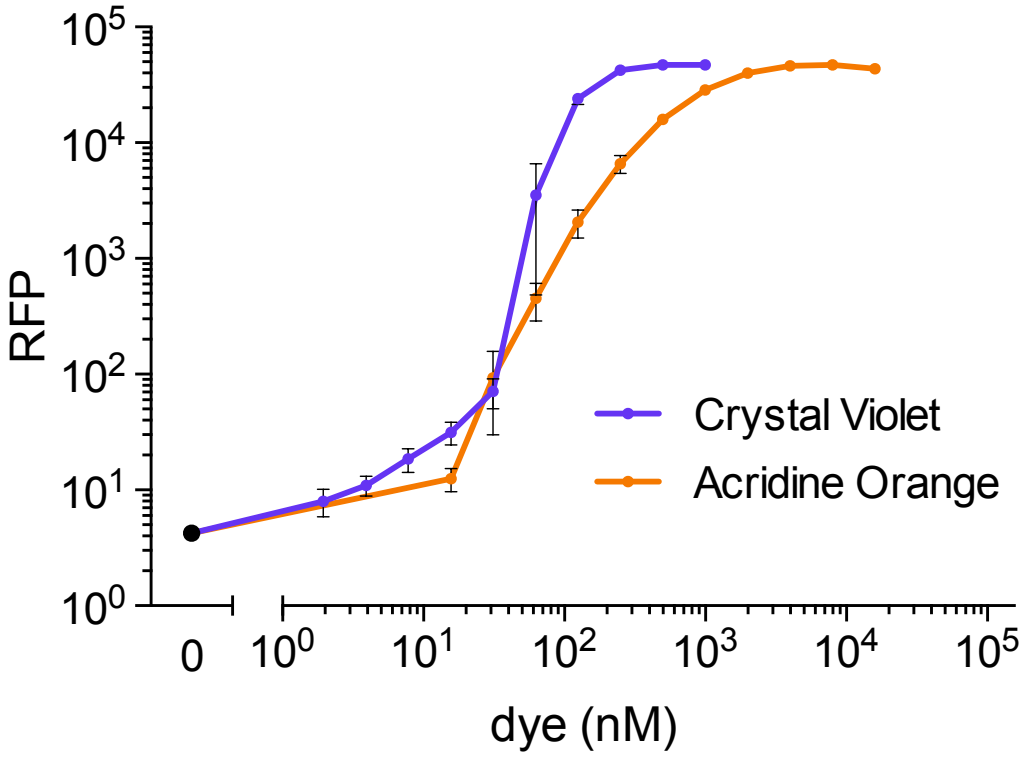


Figure 33 | Differential response of the DE20-eilOx promoter to inducing molecules. Fluorescence of *E. coli* cultures expressing RFP from the DE20-eilOx promoter that was induced either with acridine orange or crystal violet. Values and error bars represent the means and standard deviation of measurements from three independently grown cultures, which were grown to stationary phase in terrific broth.

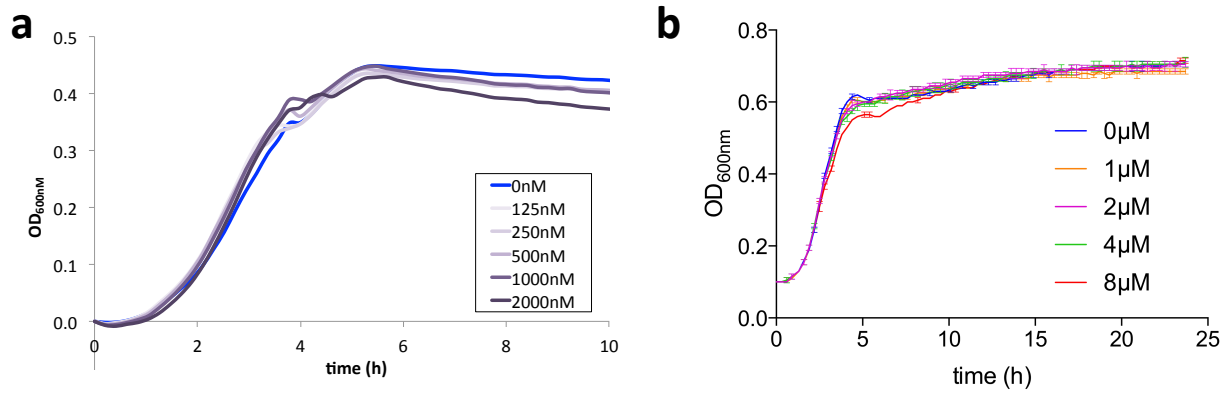
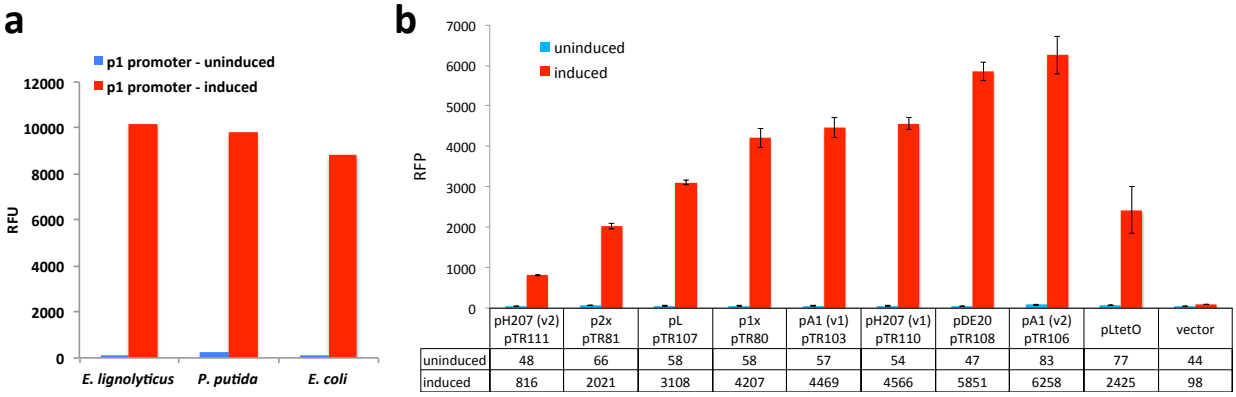


Figure 34 | Growth of *E. coli* in the presence of inducers at concentrations typically used for inducing Eil-promoters. (a) Growth of *E. coli* BL21-DE3 at 37°C in EZ-Rich defined medium containing 0.2% glucose and varying concentrations of crystal violet. Since crystal violet absorbs at 600nM, OD_{600} values were normalized to OD_{600} at start. **(b)** Growth of *E. coli* DH10B grown in EZ-Rich defined medium containing 0.2% glucose and varying concentrations of acridine orange.

3.3.1.3 Functionality of promoters in other proteobacteria

The aim of this short section was to test whether EilR can also regulate gene expression in other bacteria. Hence, the developed promoters, together with the *eilR* gene were transferred to broad host range plasmids without any modifications, and then introduced into other proteobacteria, namely the soil-dwelling microbe *Pseudomonas putida* KT2440, the nitrogen fixing bacterium *Sinorhizobium meliloti,* and *E. lignolyticus* as the original host of EilR.

On a medium to high copy number plasmid (BBR origin of replication), the "prototype" p1 promoter (see section 2.3.2 in Chapter 2), was functional in *E. lignolyticus* and *P. putida.* In these hosts, EilR-mediated repression and acridine orange-triggered induction of the p1 promoter was comparable with that of *E. coli* (Figure 35a). On a low copy plasmid (ori*V* origin of replication), the EilR-regulated phage promoters, as well as the p1x and p2x promoters (see Figure 25), were functional in *P. putida.* These promoters were tightly repressed and efficiently induced by 1μM crystal violet (Figure 35b). The same ori*V* plasmids were introduced into *S. meliloti.* Also in this bacterium, all promoters were inducible by crystal violet.



	pH207 (v2) pTR111	p2x pTR81	pL pTR107	p1x pTR80	pA1 (v1) pTR103	pH207 (v1) pTR110	pDE20 pTR108	pA1 (v2) pTR106	pLtetO	vector
uninduced	48	66	58	58	57	54	47	83	77	44
induced	816	2021	3108	4207	4469	4566	5851	6258	2425	98

Figure 35 | Eil-regulated promoters are functional in several gamma-proteobacteria. (a) Fluorescence measurements of stationary phase *Escherichia coli, Pseudomonas putida* KT2440 and the native host of EilR, *Enterobacter lignolyticus,* show that these bacteria express RFP from the EilR-regulated p1-promoter when induced by acridine orange. All strains contain the same plasmid (BBR origin of replication) and were grown in EZ-Rich medium containing 0.2% glucose. **(b)** The EilR-regulated promoters listed in **Figure 25** were transferred to *Pseudomonas putida* KT2440 via a low copy number plasmid (ori*V* origin of replication) and induced with 1μM crystal violet [1]**.** The pLtetO promoter was used as a positive control. Note that cellular background fluorescence was not deducted in (b). Values and error bars in (b) represent the means and standard deviation of measurements from three independently grown cultures, which were grown to stationary phase in LB broth.

[1] Results presented in Figure 35b and Figure 36 were obtained by Joseph Chen (San Francisco State University)

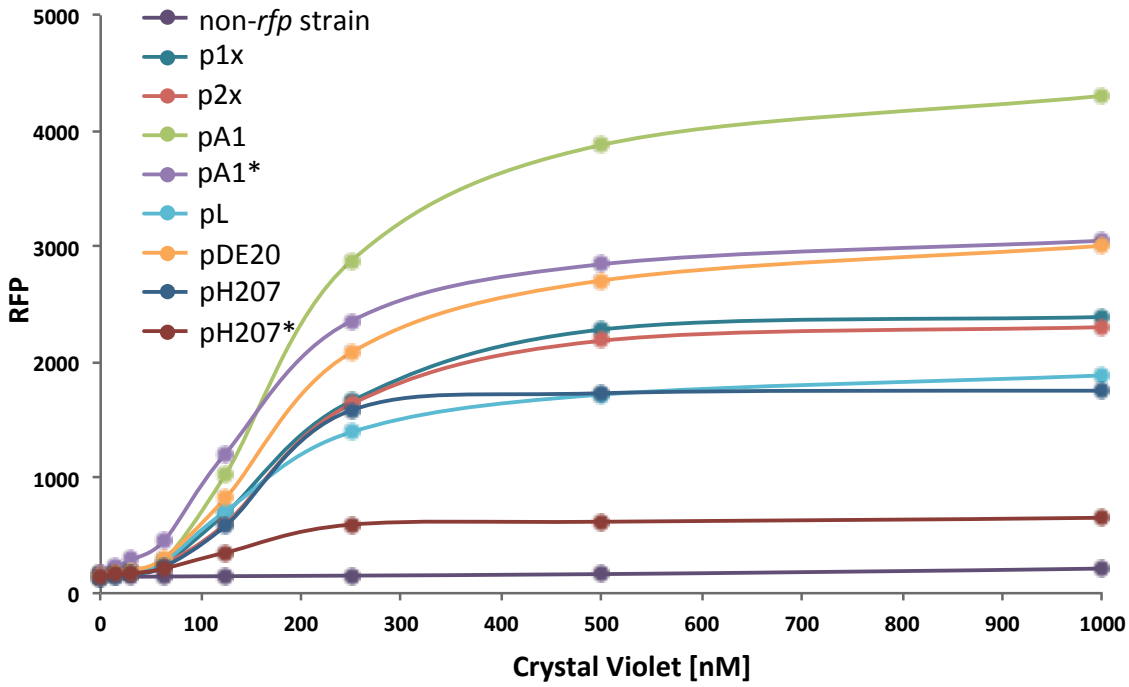


Figure 36 | EilR-mediated promoters are functional in *Sinorhizobium meliloti.* Fluorescence of *S. meliloti* expressing RFP from EilR-regulated promoters. Prior to measurements, cultures were grown to early stationary phase in PYE medium containing crystal violet at different concentrations. The asterisk indicates variants of the corresponding promoters with an eilO operator placed differently into the core promoter region.

3.3.2 EilR-mediated inducible gene expression in *S. cerevisiae*

3.3.2.1 Approach

Here, the EilR-mediated regulation mechanism is introduced into yeast with the objective to control gene expression in this eukaryotic organism. The basic principle consists of the insertion of eilO operators into a yeast promoter to enable EilR controlled transcriptional regulation.

A requirement to achieve this goal is the choice of a promoter that exhibits high and constant activity in any growth stage and media. These requirements are met by the strong constitutive TEF1 promoter (TEF1p) (Partow *et al.*, 2010), known to drive expression of a polypeptide chain elongation factor in *S. cerevisiae* (Nagashima *et al.*, 1986). For these reasons, I chose TEF1p for the insertion of eilO operators into several sites of its core promoter.

The first challenge was to find a spot in the TEF1p sequence where eilO enables repression when bound to EilR while intrinsic strength of the wild type promoter is maintained. It is important to bear in mind that slight modifications in core promoter regions can result in drastic decreases of their activity as has been shown by single base pair mutations of TEF1p (Nevoigt *et al.*, 2006). Therefore, conservation of the wild type TEF1p was taken into account when locations for eilO integration were chosen. Thus, the unconserved two base pair spacer that separates the palindromic half sites of the consensus operator was adjusted to the wild type TEF1p allowing 8 (eilO-A) to 13 base pairs (eilO-B) to be maintained. In one version (eilO-Bx), three base pairs of the consensus operator were modified to increase conservation of the wild type TEF1p to 16 base pairs (Figure 37).

Another requirement is that the EilR protein is present in the nucleus at sufficiently high levels in order to pursue its regulatory role. For this, the bacterial *eilR* gene was optimized for eukaryotic codon usage and supplied with a nuclear localization signal. The modified *eilR* gene was placed under the control of the GAL1 promoter, which is repressed by glucose and active in the presence of galactose, enabling the establishment of two scenarios to tackle important questions in a stepwise approach, as illustrated in Figure 38.

Question 1: How does eilO-insertion affect intrinsic promoter activity?

To answer this question, cells were grown in glucose-containing medium. As EilR should not be expressed under these conditions, the yellow fluorescent reporter protein (YFP) should be constitutively expressed. This scenario will give insight how insertion of the eilO operator impacts intrinsic promoter activity.

Question 2: Which site on TEF1p is suitable for repression?

In the presence of galactose, EilR is expressed and can bind to the operator located at different sites in TEF1p. This scenario will enable investigation of EilR-mediated repression and the induction by effector molecules.

a

attttatcacgtttctttttcttgaaaatttttttttgatttTTTCTCTTTCGatGACCTCCCATTgatatt
TAAGTTAATAAacGGTCTTCAATTtctCAAGTTTCAGTttCATTTTTCTTGttctattacAACT
TTTTTTActTCTTGCTCATTagaaagaaagCatagcaatctaatctaagttttctagaaaaATG

b

TEF1 site A	TTTCTCTTTCGatGACCTCCCATT	8 bp identity	1 TTTCTCTTTCGatGACCTCCCATT 24 1 AAAGTTGGACAatTGTCCAACTTT 24
eilO-A	AAAGTTGGACAatTGTCCAACTTT		
TEF1 site B	TAAGTTAATAAacGGTCTTCAATT	13 bp identity	1 TAAGTTAATAAacGGTCTTCAATT 24 1 AAAGTTGGACAacTGTCCAACTTT 24
eilO-B	AAAGTTGGACAacTGTCCAACTTT		
TEF1 site B	TAAGTTAATAAacGGTCTTCAATT	16 bp identity	1 TAAGTTAATAAacGGTCTTCAATT 24 1 TAAGTTGAACAacTGTCCAACATT 24
eilO-Bx	TAAGTTGAACAacTGTCCAACATT		
TEF1 site C	CAAGTTTCAGTttCATTTTTCTTG	12 bp identity	1 CAAGTTTCAGTttCATTTTTCTTG 24 1 AAAGTTGGACAttTGTCCAACTTT 24
eilO-C	AAAGTTGGACAttTGTCCAACTTT		
TEF1 site D	AACTTTTTTTActTCTTGCTCATT	12 bp identity	1 AACTTTTTTTActTCTTGCTCATT 24 1 AAAGTTGGACActTGTCCAACTTT 24
eilO-D	AAAGTTGGACActTGTCCAACTTT		

Figure 37 | Insertion of single 24bp eilO operators into the TEF1 core promoter. **(a)** Sequence of the native yeast TEF1 core promoter region showing the four sites (blue, brown, green, orange) that were replaced with eilO operators. The transcriptional start is indicated in red, and the translational start in pink. The predicted TATA box-like region is underlined.
(b) Sequences of the eilO versions (in black) aligned to their corresponding sites in the native TEF1 promoter, with the number of matching base pairs listed. The two non-conserved central base pairs that separate the operator half-sites are shown in lower case letters. Base pairs that deviate from the consensus operator are shown in pink. In the right column, the same sequences are depicted in a way that shows the mutations caused by eilO insertion (shown in red).

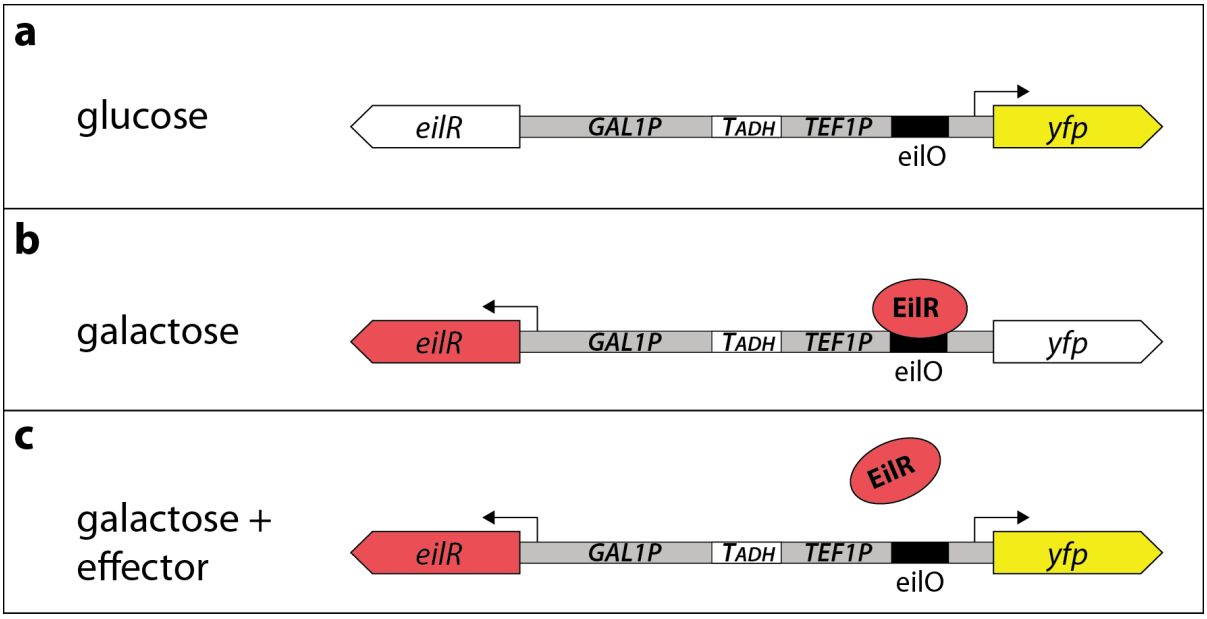


Figure 38 | Stepwise approach to investigate functionality of EilR-mediated transcription in yeast. On a single-copy plasmid, the *eilR*-gene is under control of the GAL1 promoter (GAL1p). Separated by a transcriptional terminator (T_{ADH}), the TEF1 promoter (TEF1p) containing the eil-operator (eilO) at different sites drives expression of *yfp,* the reporter gene encoding YFP. Arrows indicate the active state of the promoter. Genes are marked in color when expressed, and shown in white when repressed. **(a)** In the presence of glucose, GAL1p is repressed, preventing expression of the EilR repressor. As a result, *yfp* is constitutively transcribed from TEF1p, allowing inspection of how eilO insertion affects intrinsic TEF1p activity. **(b)** Galactose triggers expression of EilR, which now binds to the operator sites located in TEF1p to repress *yfp*. This scenario enables the examination how the site at which eilO is inserted into TEF1p influences EilR-mediated repression. **(c)** In the presence of effector molecules and the repressor, the capability of these EilR-ligands to induce YFP expression can be measured.

3.3.2.2 EilR and its ligand regulate the TEF1 promoter

Insertion of an eilO operator decreased intrinsic promoter activity, which was expected since few mutations can already severely reduce TEF1p activity, as mentioned above. The smallest decrease of activity was observed with operator eilO-Bx (Figure 39).

Operators eilO-Bx & eilO-B, both being located at the same site in TEF1p, overlap the predicted TATA-like area. Insertion of the consensus operator eilO-B, which maintains only 11 base pairs of the original wild type TEF1p sequence, resulted in a 3-fold decrease of promoter activity compared to the wild type. The eilO-Bx version deviates from the consensus operator by three base pairs to increase conservation of the wild type TEF1p to 16 base pairs. Activity of the TEF1-eilO-Bx promoter is much higher than the one of eilO-B and nearly reached activity level of the wild type promoter in the absence of EilR. When EilR was expressed, eilO-Bx was only slightly

"leakier" than eilO-B, but it is unclear whether the reason for this was caused by weaker EilR binding or by higher intrinsic activity. It is noteworthy that YFP levels expressed from the *eilR* containing TEF-wt plasmid was 5 to 6 times lower compared to the TEF-wt* reference plasmid (Figure 39)(Alper *et al.*, 2005). This is unlikely caused by unspecific binding of EilR, since a reduced signal was observed in the absence and the presence of the repressor. A possible reason could be the lack of a transcriptional terminator for the *yfp* gene in all developed plasmids (Curran *et al.*, 2013).

The data indicate that EilR protein is functional in yeast and present at levels that are sufficient for repression. The ability of EilR to repress TEF1 is determined by the localization of its binding site in the promoter region. Highest repression was achieved when the operator either overlapped with the predicted TATA-like element (Nagashima *et al.*, 1986) (eilO-B & eilO-Bx) or when located in its 3' proximity (eilO-C). EilR-binding in these regions might hinder transcription factor binding, therefore inhibiting recruitment of RNA-polymerase. Repression was weaker when the operator was placed either upstream of the TATA box-like region or at the distal 3' end of the promoter.

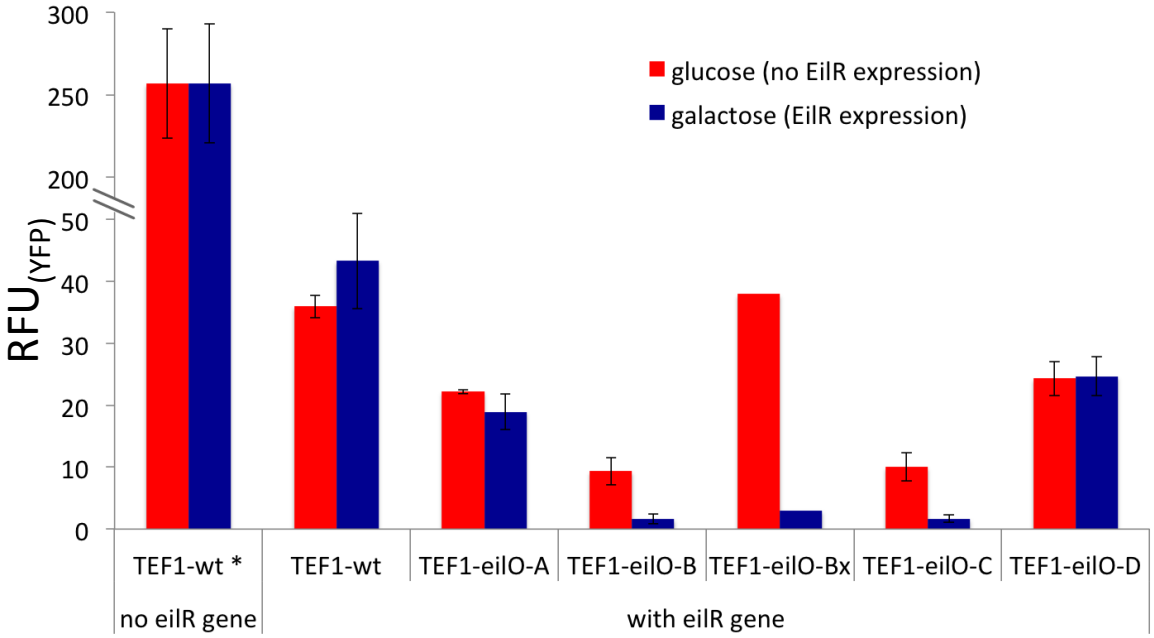


Figure 39 | EilR-mediated repression is dependent on localization of the eilO operator in the TEF1 promoter. Yeast cells expressing YFP from the wild type promoter (TEF1-wt) or from modified promoters containing one eilO operator at different sites (TEF1-eilO-A to TEF1-eilO-D). After 20h of growth, EilR (driven from the GAL1 promoter) is not expressed in media containing 2% glucose (red), but in 2% galactose (blue). TEF1-wt* is the reference plasmid (Alper *et al.*, 2005) that does not contain the GAL1-driven *eilR* gene. With the

exception of TEF1-eilO-Bx, values and error bars represent the means and standard deviation of measurements from four independently grown colonies.

These observations illustrate that it is very delicate to find a spot in the TEF1 promoter sequence where eilO enables tight repression when bound to EilR while maintaining intrinsic strength of the wild type promoter.

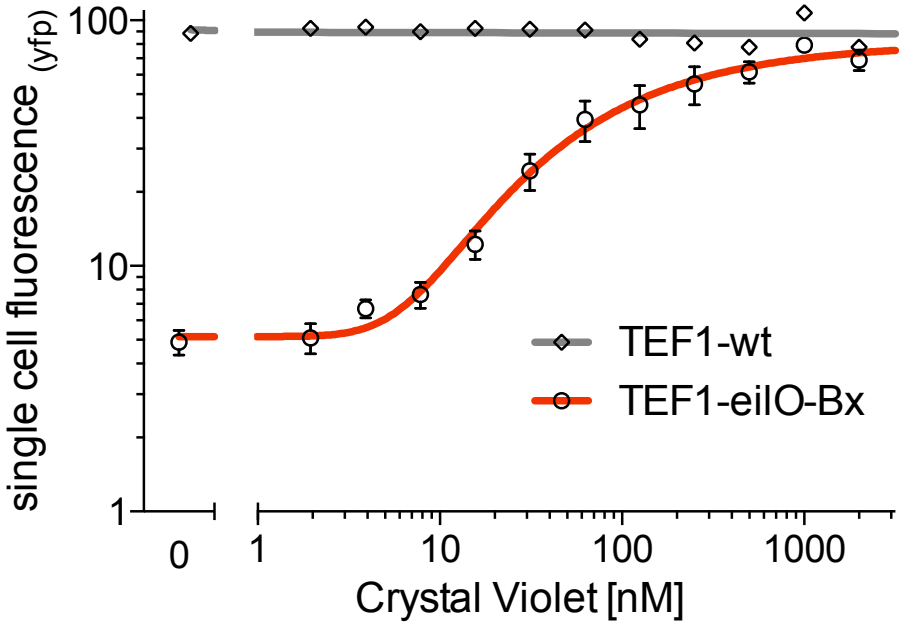


Figure 40 | Crystal violet de-represses the TEF1-eilO-Bx promoter. Yeast cells expressing YFP from the EilR-regulated TEF1-eilO-Bx promoter or the wild type promoter (TEF1-wt) grown in different concentrations of crystal violet. Crystal violet de-represses the TEF1-eilO-Bx but does not affect the wild type promoter. Cultures were grown for 20 hours; crystal violet was added at the time of inoculation. Values and error bars represent the means and standard deviation of measurements from two independently grown cultures. Background fluorescence was subtracted from a non-YFP strain**.**

Out of the five tested promoter variants, TEF1-eilO-Bx showed the highest ON/OFF rate when switching from glucose to galactose. Therefore, this version was selected for experiments with crystal violet as effector molecule. Single cell fluorescence measurements revealed that activity of this promoter shows a clear concentration-dependent response. De-repression caused by 1μM crystal violet resulted in an approximately 30-fold increase of YFP expression, reaching levels similar to the ones of the wild-type promoter.

3.4 Conclusions & Outlook

In the context of this thesis, I discovered and characterized both the EilR as negative regulator of multidrug efflux as well as its cognate operator, eilO. Phylogenetic analysis of homologous multidrug efflux systems enabled the generation of a consensus operator with an increased affinity for the EilR repressor protein. This finding enabled the development of bacterial promoters that are inducible over more than four orders of magnitude. The developed promoters comprise many desirable features, including tight repression in the absence of effector molecules and high transcription rates when induced. Despite the ligand-multispecificity of EilR, this repressor provides orthogonal transcriptional regulation, without being affected by media conditions or other expression systems.

Although the developed bacterial promoters were optimized for functionality in *E. coli,* they also operate in *S. meliloti* and *P. putida*. I expect that EilR-mediated transcriptional regulation in these and other bacteria can be further improved by combining the eil-operator with host-specific promoters.

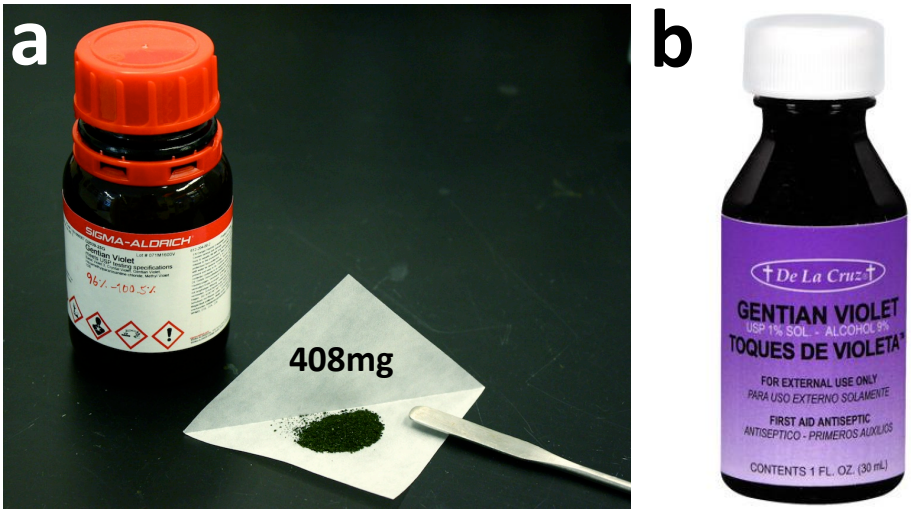


Figure 41 | Crystal violet as an effective and readily available inducer molecule. (a) 408 mg of crystal violet (= gentian violet) are sufficient to induce a 1000 Liter fermentation reaction at a 1µM concentration. **(b)** Crystal violet has a wide range of applications. It is used as stain to identify gram-positive bacteria, as an ink component for pens and printers, as a topical antiseptic agent and as dye to colorize hair, leather or silk (FAO-WHO, 2013; National Biochemicals Corp., 2013). As shown in the figure, this inducer molecule is also sold in department stores and pharmacies as topical antifungal agent.

My results of experiments with *S. cerevisiae* demonstrate that the codon-optimized EilR binds to its operator located on a eukaryotic promoter, resulting in repression of the reporter gene. Inducer molecules such as crystal violet de-repress this promoter and enable regulated, concentration-dependent gene expression. Compared to the EilR-regulated bacterial promoters, the developed yeast promoter is still very leaky. While it nearly reached the activity of the wild-type promoter when induced, basal activity was still considerable in the uninduced state. Analogously to the developed bacterial promoters, insertion of additional operators into candidate sites should increase the dynamic range. Taken together, the experiments in *S. cerevisiae* indicate that EilR-mediated regulation has the potential to be expanded to other eukaryotic organisms for orthogonal gene expression.

Table 4 | List of commonly used inducible expression systems, with the typical concentration of inducer for full induction and the respective costs to induce a 1000L fermentation.

Host	Promoter	Inducer	Typical concentration for full induction	Cost per 1000L (USD) *	Reference
E. coli	*LAC*	Isopropyl β-D-1-thiogalactopyranoside	200 μM	2000.00	(Studier & Moffatt, 1986)
E. coli	*araBAD*	Arabinose	13 mM (0.2%)	1578.00	(Guzman et al., 1995)
E. coli	*rhaBAD*	Rhamnose	12 mM (0.2%)	285.00	(Haldimann et al., 1998)
S. cerevisiae	*GAL1*	Galactose	111 mM (2%)	2464.00	(Partow et al., 2010)
S. cerevisiae	*CUP1*	Copper sulfate	1 mM	10.00	(Hottiger et al., 1994)
E. coli	*TET*	Anhydrotetracycline	400 nM	308.00	(Skerra, 1994)
E. coli	*CYM*	Cumate	100 μM	107.00	(Choi et al., 2010)
E. coli	*EIL*	Crystal violet	1 μM	0.15	
E. coli	*EIL*	Acridine orange	10 μM	12.00	

* prices are based on the least expensive unit sold at Sigma-Aldrich (April 2015)

The molecules needed to induce EilR-mediated promoters are potent, widely used (Figure 41) and extremely inexpensive. The price for induction by crystal violet is at least 1000 times lower than the one of existing inducible gene expression systems commonly used for large-scale fermentations (Table 4). An exception is the leaky copper-inducible system in yeast (Hottiger *et al.*, 1994), which is still 100 times more expensive than the promoters I developed. Particularly in large-scale production of enzymes, biofuels and bulk chemicals, the cost of current inducer compounds may prohibit the use of a preferred expression system at the expense of complete external control, choice of media and maximal yield (Keasling, 2012). EilR-mediated promoters render such tradeoffs unnecessary and could thus provide a way for tightly regulated high-level gene expression in any application and host organism.

3.5 Experimental procedures

Several procedures applied in this chapter are described in Chapter 2.

3.5.1 Bacteria

Construction of promoters

An overview of the DNA parts used for promoter construction is listed in Table 5.

Phage promoters

Sequences for wild-type phage promoters were retrieved from Deuschle et al., 1986. Operators were placed into The 17bp wild type spacer region of the phage promoters was replaced with the truncated consensus operator in a way that at least one operator basepair overlapped either the -35 or -10 site. Designed DNA sequences were ordered as gBlocks (IDT), with the flanking regions containing at least 30 basepairs identity with the backbone ends. As vector backbone, modified p1 biosensor plasmid was used (Chapter 2). Modifications include replacement of the bicistronic 5' UTR with the following sequence obtained from (T. S. Lee *et al.*, 2011) (underlined):
(p1 -10 site)…gtgtggaGAATTCAAAAGATCTTTTAAGAAGGAGATATACAT...(start *rfp*). gBlocks were cloned into the linearized vector backbone by isothermal DNA assembly (Gibson *et al.*, 2009), following the manufacturer's (New England Biolabs) instructions. The resulting plasmids (Table 5a) containing promoters with one operator were transformed into *E. coli* DH10B, sequence verified and used as template for insertion of the 2nd operator downstream of the -10 site. Plasmids were PCR-amplified with primers that each contained half an operator. PCR products were self-ligated, transformed into *E. coli* DH10B and sequence verified (Table 5b).

p1x and p2x promoters

As described in the phage promoter section, modified p1 or p2 sensor plasmids were used as templates for PCR. Backbones were amplified with primers listed in Table 5b, self-ligated, transformed into *E. coli* DH10B and sequence verified.

ig-eiAR-v1/v2 promoters

In a first step, the *eilR* gene and the 117 basepairs intergenic region were PCR amplified from an IL-tolerant fosmid (Chapter 1) with primers containing a ca. 20 basepair overlap with the backbone ends. As backbone, linearized pFAB5088 was used (Mutalik *et al.*, 2013), and DNA parts were linked by isothermal DNA assembly. After transformation and sequence verification, the resulting plasmid (ig-eilAR-wt)

was used as template for inserting the mutations to generate modified promoters ig-eilAR-v1/v2 (Table 5b).

Table 5 | Sequences of primers and DNA-parts used for promoter construction. (-35 and -10 sites are shown in upper case letters.)

a

	promoter sequence incl. operator (variable part of ordered gBlock)		resulting promoter name (one operator)
5' backbone (p1 biosensor)	aaaatttatcaaaaagagtaTTGACTtggacacgtgtccaactGATACTtacagccAtcgagagggacacggcga	3' backbone (p1 biosensor)	pA1-eilO_v1
	aaaatttatcaaaaagagtaTTGACTtaaagttggacacgtgtGATACTtacagccAtcgagagggacacggcga		pA1-eilO_v2
	aaaaaactgcaaaaaatagtTTGACAggacacgtgtccaacttTAAGATgtacccAgttcgatgagagcgataac		pDE20-eilO
	ataatttttaaaaaattcatTTGCTAaagttggacacgtgtccTATAATatacttcAtaaattgataaacaaaaa		pH207-eilO_v1
	ataatttttaaaaaattcatTTGCTAggacacgtgtccaacttTATAATatacttcAtaaattgataaacaaaaa		pH207-eilO_v2
	ataaattatctctggcggtgTTGACAaagttggacacgtgtccGATACTgagcacAtcagcaggacgcactgacc		PL-eilO

b

reverse primer name	reverse primer sequence	annealing temp (°C)	F-primer used	resulting promoter name (two operators)	resulting promoter #
pA1-eilO_v1-R	gtgtccaacttTggctgtaAGTATCagttgg	59	Phage_High_F	pA1A-eilOx	pTR103
pA1-eilO_v2-R	gtgtccaactttggctgtaAGTATCacacg	59	Phage_High_F	pA1B-eilOx	pTR106
pDE20-eilO-R	gtgtccaactttgggtacATCTTAaagtt	52	Phage_Low_F	pDE20-eilOx	pTR108
pH207-eilO_v1-R	gtgtccaacttTgaagtatATTATAggaca	52	Phage_Low_F	pH207A-eilOx	pTR110
pH207-eilO_v2-R	gtgtccaacttTgaagtatATTATAaagtt	52	Phage_Low_F	pH207B-eilOx	pTR111
PL-eilO-R	gtgtccaacttTgtgctcAGTATCggaca	59	Phage_High_F	pL-eilOx	pTR107
eilO_P1x-R	GTGTCCAACTTTccacacATCATAGTTGGAC	63	Px_F	p1x	pTR80
eilO_P2x-R	GTGTCCAACTTTccacacACATTAgttggacac	65	Px_F	p2x	pTR81
IG-5'_wt_R	ACACTTGTCCAGCTTTTTGTTATTATCGCCC	63	IG-3'_mut_F	ig-eilAR-v1	pTR85
IG-5'_mut_R	CACTTGTTCAACTTTTTGTTATTATCGCCCTATCACC	63	IG-3'_mut_F	ig-eilAR-v2	pTR86

forward primer name	forward primer sequence
Phage_Low_F	gtgtccaactttGAATTCAAAAGATCTTT
Phage_High_F	gtgtccaactttGAATTCAAAAGATCTTTTAAGAAGGAG
Px_F	GTGTCCAACTTTgaattcaaaagatctTTTAAGAAGGAGATATACATatg
IG-3'_mut_F	TTCAACTTTCCCTATAATCGCAAACTGGACG

Experiments with P. putida, E. lignolyticus and S. meliloti

BBR-plasmid (*P. putida* and *E. lignolyticus*)

The p1 promoter, including the *eilR* gene (see Chapter 2), was inserted to pbbB7k (kanR, BBR origin of replication) (T. S. Lee *et al.*, 2011). The plasmid was electroporated into *P. putida* and *E. lignolyticus* following the instructions described in Diver *et al.*, 1990. Both strains were grown at 30°C in LB medium containing 50mg/L of kanamycin.

oriV-plasmid (*P. putida* and *S. meliloti*)

The p1x, p2x and developed phage promoters including the *eilR* gene were inserted without modifications into a broad host range plasmid described in Marx & Lidstrom, 2001. Mobilization of plasmids from *E. coli* to *S. meliloti* and *P. putida* was accomplished by triparental mating, with the help of strain HB101, carrying pRK2073

(*P. putida)* or MT616, carrying pRK600 (*S. meliloti)* (Fields *et al.*, 2012). *P. putida* were grown at 30°C in LB medium containing 50mg/L of kanamycin. *S. meliloti* were grown at 30°C in PYE medium containing 25mg/L of kanamycin.

3.5.2 Yeast

As backbone, I used a single-copy centromeric plasmid driving expression of the reporter gene encoding the yellow fluorescent protein yEcitrine (here named YFP*)* from the TEF1 promoter (Alper *et al.*, 2005).

The first attempt to create an EilR-regulated TEF1p did not give any interpretable results. The "updated" version (Figure 42) contained the following changes:

- The *eilR* gene was optimized for eukaryotic codon usage
- A nuclear localization signal (GlyProLysLysLysArgLysVal) was added to the C-terminal of EilR to enhance transport from the cytosol to the nucleus.
- The divergent GAL1 and TEF1 promoters were isolated from each other with the ADH1 transcriptional terminator to avoid crosstalk.
- The CYC transcriptional terminator was placed after the *eilR* gene.

DNA harboring the parts T-CYC, *eilR,* GAL1p, T-ADH1, TEF1p (wild type) and the *yfp* gene were de-novo synthesized (Genscript, Piscataway, NJ). Unique restriction sites separate these parts from each other for convenient exchange of components and reassembly. The synthesized DNA was then cloned into the KpnI/SacI restriction sites of pRS416, a single-copy centromeric shuttle vector containing the URA3 marker gene (Sikorski & Hieter, 1989). The wild type TEF1p on the resulting plasmid pRS416_TEFwt_eilR was then exchanged by restriction cloning at the XhoI/XbaI restriction sites with the de-novo synthesized TEF1 variants containing the eilO operators (Figure 37).

The plasmids were transformed into *S. cerevisiae* strain BY4741 (URA^-: uracil auxotroph) and grown in defined, uracil-deficient medium containing 2% glucose. For EilR-binding assays, cells were grown in uracil-deficient medium containing either 2% glucose or 2% galactose. The empty plasmid pRS416 was used as negative control. A previously published plasmid (Alper *et al.*, 2005) that constitutively drives *yfp* expression from the wild type TEF1p, was used as a positive reference.

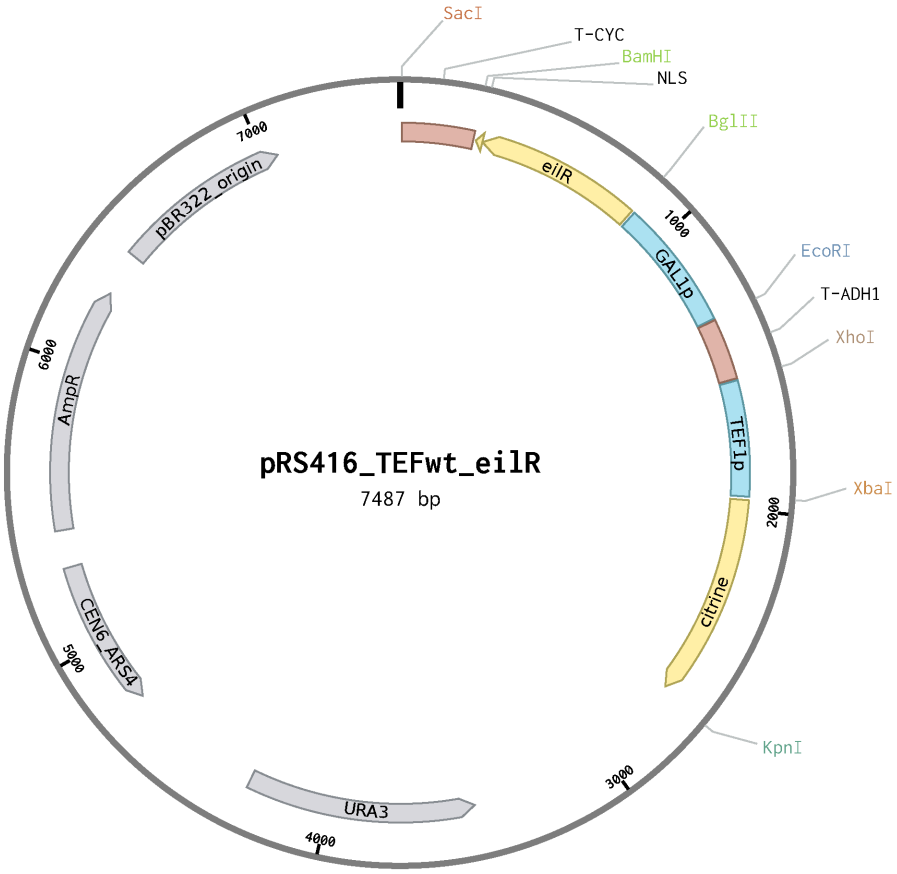


Figure 42 | Map of plasmid used for the insertion of eilO operators. Elements relevant for this experiment are shown in color. Legend: transcriptional terminators (red); promoters (blue); genes (yellow); NLS=nuclear localization signal.

CONCLUDING REMARKS

The results of my research presented in this thesis demonstrate how environmental microorganisms provide an inexhaustible resource for uncovering unknown functions. The discovery of a gene pair from a rain forest bacterium marked the beginning of a broad scientific journey, during which I gained insight on how bacteria possess an elaborate toolset to respond to novel stressors. The Eil efflux system provides an example of how novel functions can be converted into biotechnological applications, either by specifically targeting an objective (IL tolerance), or after engagement with the discovery (expression system).

The functional genetic screening approach taken in Chapter 1 proved to be a powerful method for discovering the mechanisms underpinning phenotypes such as bacterial resistance to inhibitory xenobiotic compounds. In future studies, this sequence-independent method could be expanded to detect other useful functions, such as bacterial enzymes that degrade lignin. Such enzymes will be useful in biofuel applications, because lignin makes the largest contribution to recalcitrance in biomass. Moreover, the aromatic lignin monomer products are valuable resources, for example to produce biopolymers (Linger *et al.*, 2014), or lignin-derived ionic liquids (Socha *et al.*, 2014).

Expanding the screening approach from the genome of a single microbial isolate to metagenomes from environmental samples will enable access to functional genes in uncultivable organisms (Handelsman *et al.*, 1998). Such metagenomic libraries store an immense amount of genetic information and allow for the screening of multiple phenotypes. Rather than performing a single experiment for each metagenomic library, the idea of “open resource metagenomics” (Neufeld *et al.*, 2011) promotes sharing these libraries among many researchers in order to realize the potential of this *physically* available genetic information.

The option to screen DNA-libraries in multiple host organisms increases the chance of uncovering a desired phenotype. For this reason, the DNA is best stored in a standardized system that can easily be transferred to other hosts and vectors (Neufeld *et al.*, 2011; Troeschel *et al.*, 2010). Ideally, the screening system also enables promiscuous transcription, for example by host-independent RNA polymerases (Arvani *et al.*, 2012; Terrón-González *et al.*, 2013) see also functional screening discussion, p. 28.

The challenge of functional screening methods is well illustrated by the experiments conducted with methyl viologen. Since EilR is not sufficiently responsive to this inhibitor, it prevents EilA-mediated establishment of tolerance. However, it is very likely that exposure of *E. lignolyticus* to methyl viologen over an extended period of time would evolve a tolerant strain. As I have shown in Chapter 3, mutating three base pairs in the native *eilA* promoter sequence resulted in a strong increase in promoter activity. Selective pressure in the presence of an inhibitor can cause the same type of mutations (Blair *et al.*, 2014; Nikaido, 2009), providing a simple way to adapt to new stresses. Cumulative stress of a toxic substrate and pump overexpression can be minimized by repressor-mediated negative autoregulation. Such a mechanism provides a very sensitive means to appropriately adjust pump levels to inhibitor concentrations, as was observed with the *E. lignolyticus* and *S. typhimurium* system in response to ILs.

All of the identified EilR ligands are molecules not encountered in the native environment of *E. lignolyticus.* Thus, major questions regarding the Eil efflux system are: what is the natural substrate(s), and is its function in the natural environment. Does this efflux system export an endogenously produced metabolite? Or does it respond to environmental signals released by other organisms? The activity guided approach described in Chapter 2, as well as co-culturing approaches with bacterial communities (Shank *et al.*, 2011) could provide ways to uncover natural substrates of MDR efflux systems.

As a multidrug binding protein, EilR responds to a fairly wide range of structurally dissimilar compounds. This multispecificity contrasts with transcription factors that regulate other expression systems, which are defined by their very narrow substrate specificity. Yet, the developed EilR-mediated gene expression systems operate orthogonally, without being affected by media conditions or other expression systems.

The approach taken in Chapter 3 to develop optimized EilR-mediated promoters can be translated to tailor tunable promoters for specific applications. A possible way to identify transcription factors that respond to a small molecule of choice is described by Uchiyama *et al.*, 2005, which uses a reporter-gene coupled approach to screen metagenomic DNA. Otherwise, TetR family repressors with known effector molecules are suitable candidates for the development of an induction system, due to the direct mode of regulation and the ability to achieve high specificity with relatively short operator sequences (Stanton *et al.*, 2014). As was done with eilO, binding affinity of the operator to the selected transcription factor can be increased

by phylogenetic analysis prior to its insertion into a strong promoter. Induction systems based on phage promoters, which are specialized for efficient transcription in *E. coli,* have been widely used in other bacterial hosts (Gamer *et al.*, 2009; Lussier *et al.*, 2010; Wang *et al.*, 2000, and this work). However, to engineer an inducible expression system for in non *E. coli* hosts, one should focus more on promoters from numerous host-specific phages, which are optimized by nature to operate efficiently in the selected bacterium.

The developed promoters comprise several desirable features, including tight repression, high transcription rates in *E. coli* when induced, media independency, and absence of interference with other induction systems. In addition, the negligible costs of inducer molecules render any compromise between costs and controllability unnecessary, which opens the way for tightly regulated high-level gene expression in any scale and host organism, including eukaryotes. Referring to its origin, I named this colorful way to control gene expression “Jungle Express”.

El Yunque cloud forest in Puerto Rico, where *Enterobacter lignolyticus* was isolated (see also cover image). I had the chance to visit this magical spot during a sampling field trip.

There is an endless Jungle full of novel functions waiting to be discovered.

It is rewarding to join the expedition…

BIBLIOGRAPHY

Ahmed, M., Lyass, L., Markham, P. N., Taylor, S. S., Vazquezlaslop, N., & Neyfakh, A. A. (1995). 2 Highly Similar Multidrug Transporters of Bacillus-Subtilis Whose Expression Is Differentially Regulated. *Journal of Bacteriology*, *177*, 3904–3910.

Ahn, S. K., Cuthbertson, L., & Nodwell, J. R. (2012). Genome Context as a Predictive Tool for Identifying Regulatory Targets of the TetR Family Transcriptional Regulators. *PLoS ONE*, *7*.

Ajikumar, P. K., Xiao, W.-H., Tyo, K. E. J., Wang, Y., Simeon, F., Leonard, E., Mucha, O., Phon, T. H., … Stephanopoulos, G. (2010). Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli. *Science*, *330*, 70–74.

Alper, H., Fischer, C., Nevoigt, E., & Stephanopoulos, G. (2005). Tuning genetic control through promoter engineering. *Proc. Natl. Acad. Sci.*, *102*, 12678–12683.

Arvani, S., Markert, A., Loeschcke, A., Jaeger, K. E., & Drepper, T. (2012). A T7 RNA polymerase-based toolkit for the concerted expression of clustered genes. *Journal of Biotechnology*, *159*, 162–171.

Bailey, T. L., & Elkan, C. (1994). Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *ISMB. International Conference on Intelligent Systems for Molecular Biology*, *2*, 28–36.

Balzer, S., Kucharova, V., Megerle, J., Lale, R., Brautaset, T., & Valla, S. (2013a). A comparative analysis of the properties of regulated promoter systems commonly used for recombinant gene expression in Escherichia coli. *Microbial Cell Factories*, *12*, 26.

Balzer, S., Kucharova, V., Megerle, J., Lale, R., Brautaset, T., & Valla, S. (2013b). A comparative analysis of the properties of regulated promoter systems commonly used for recombinant gene expression in Escherichia coli. *Microbial Cell Factories*, *12*, 26.

Baranyi, J., & Roberts, T. A. (1994). A Dynamic Approach to Predicting Bacterial-Growth in Food. *International Journal of Food Microbiology*, *23*, 277–294.

Bernard, H. U., Remaut, E., Hershfield, M. V, Das, H. K., Helinski, D. R., Yanofsky, C., & Franklin, N. (1979). Construction of plasmid cloning vehicles that promote gene expression from the bacteriophage lambda pL promoter. *Gene*, *5*, 59–76.

Bibb, M. J. (2005). Regulation of secondary metabolism in streptomycetes. *Current Opinion in Microbiology*, *8*, 208–215.

Blair, J. M. A., Webber, M. A., Baylay, A. J., Ogbolu, D. O., & Piddock, L. J. V. (2014). Molecular mechanisms of antibiotic resistance. *Nature Reviews Microbiology*, *13*, 42–51.

Blanch, H. W., Simmons, B. A., & Klein-Marcuschamer, D. (2011). Biomass deconstruction to sugars. *Biotechnology Journal*, *6*, 1086–1102.

Blazeck, J., Garg, R., Reed, B., & Alper, H. S. (2012). Controlling promoter strength and regulation in Saccharomyces cerevisiae using synthetic hybrid promoters. *Biotechnology and Bioengineering*, *109*, 2884–2895.

Bobleter, O., Niesner, R., & Röhr, M. (1976). The hydrothermal degradation of cellulosic matter to sugars and their fermentative conversion to protein. *Journal of Applied Polymer Science*, *20*, 2083–2093.

Bochner, B. R. (2003). New technologies to assess genotype-phenotype relationships. *Nature Reviews. Genetics*, *4*, 309–314.

Bokinsky, G., Peralta-Yahya, P. P., George, A., Holmes, B. M., Steen, E. J., Dietrich, J., Soon Lee, T., Tullman-Ercek, D., … Keasling, J. D. (2011). Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered Escherichia coli. *Proc. Natl. Acad. Sci.*, 1–6.

Bouvet, P., & Belasco, J. G. (1992). Control of RNase E-mediated RNA degradation by 5'-terminal base pairing in E. coli. *Nature*, *360*, 488–91.

Brandt, A., Gräsvik, J., Hallett, J. P., & Welton, T. (2013). Deconstruction of lignocellulosic biomass with ionic liquids. *Green Chemistry*, *15*, 550–583.

Brown, S., Clastre, M., Courdavault, V., & O'Connor, S. E. (2015). De novo production of the plant-derived alkaloid strictosidine in yeast. *Proceedings of the National Academy of Sciences*, 201423555.

Campbell, E. A., Korzheva, N., Mustaev, A., Murakami, K., Nair, S., Goldfarb, A., & Darst, S. A. (2001). Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell*, *104*, 901–912.

Carlioz, A., & Touati, D. (1986). Isolation of superoxide dismutase mutants in Escherichia coli: is superoxide dismutase necessary for aerobic life? *The EMBO Journal*, *5*, 623–630.

Cartwright, C. P., Li, Y., Zhu, Y. S., Kang, Y. S., & Tipper, D. J. (1994). Use of β-lactamase as a secreted reporter of promoter function in yeast. *Yeast*, *10*, 497–508.

Chen, Y., Smanski, M. J., & Shen, B. (2010). Improvement of secondary metabolite production in Streptomyces by manipulating pathway regulation. *Applied Microbiology and Biotechnology*, *86*, 19–25.

Cho, Y. H., Kim, E. J., Chung, H. J., Choi, J. H., Chater, K. F., Ahn, B. E., Shin, J. H., & Roe, J. H. (2003). The pqrAB operon is responsible for paraquat resistance in Streptomyces coelicolor. *Journal of Bacteriology*, *185*, 6756–6763.

Cohn, C. A., Pak, A., Strongin, D., & Schoonen, M. A. (2005). Quantifying hydrogen peroxide in iron-containing solutions using leuco crystal violet. *Geochemical Transactions*, *6*, 47–51.

Curran, K. A., Karim, A. S., Gupta, A., & Alper, H. S. (2013). Use of expression-enhancing terminators in Saccharomyces cerevisiae to increase mRNA half-life and improve gene expression control for metabolic engineering applications. *Metabolic Engineering*, *19*, 88–97.

Dale, B. E., & Moreira, M. J. (1982). Freeze-explosion technique for increasing cellulose hydrolysis. *Biotechnol. Bioeng. Symp.; (United States)*, *12*.

Dashtban, M., Schraft, H., Syed, T. a., & Qin, W. (2010). Fungal biodegradation and enzymatic modification of lignin. *International Journal of Biochemistry and Molecular Biology*, *1*, 36–50.

De Jong, B., Siewers, V., & Nielsen, J. (2012). Systems biology of yeast: enabling technology for development of cell factories for production of advanced biofuels. *Current Opinion in Biotechnology*, *23*, 624–630.

DeAngelis, K. M., D'Haeseleer, P., Chivian, D., Fortney, J. L., Khudyakov, J., Simmons, B., Woo, H., Arkin, A. P., … Hazen, T. C. (2011). Complete genome sequence of "Enterobacter lignolyticus" SCF1. *Standards in Genomic Sciences*, *5*, 69–85.

Dehal, P. S., Joachimiak, M. P., Price, M. N., Bates, J. T., Baumohl, J. K., Chivian, D., Friedland, G. D., Huang, K. H., … Arkin, A. P. (2009). MicrobesOnline: An integrated portal for comparative and functional genomics. *Nucleic Acids Research*, *38*.

Deuschle, U., Kammerer, W., Gentz, R., & Bujard, H. (1986). Promoters of Escherichia coli: a hierarchy of in vivo strength indicates alternate structures. *The EMBO Journal*, *5*, 2987–2994.

Diver, J. D., Bryan, L. E., & Sokol, P. a. (1990). Transformation of Pseudomonas aeruginosa by electroporation. *Analytical Biochemistry*, *189*, 75–79.

Dunlop, M. J., Dossani, Z. Y., Szmidt, H. L., Chu, H. C., Lee, T. S., Keasling, J. D., Hadi, M. Z., & Mukhopadhyay, A. (2011). Engineering microbial biofuel tolerance and export using efflux pumps. *Molecular Systems Biology*, *7*, 487.

Emory, S. a., Bouvet, P., & Belasco, J. G. (1992). A 5′-terminal stem-loop structure can stabilize mRNA in Escherichia coli. *Genes and Development*, *6*, 135–148.

Fields, A. T., Navarrete, C. S., Zare, A. Z., Huang, Z., Mostafavi, M., Lewis, J. C., Rezaeihaghighi, Y., Brezler, B. J., … Chen, J. C. (2012). The conserved polarity factor PodJ1 impacts multiple cell envelope-associated functions in Sinorhizobium meliloti. *Molecular Microbiology*, *84*, 892–920.

Fluman, N., & Bibi, E. (2009). Bacterial multidrug transport through the lens of the major facilitator superfamily. *Biochimica Et Biophysica Acta-Proteins and Proteomics*, *1794*, 738–747.

Gamer, M., Fröde, D., Biedendieck, R., Stammen, S., & Jahn, D. (2009). A T7 RNA polymerase-dependent gene expression system for Bacillus megaterium. *Applied Microbiology and Biotechnology*, *82*, 1195–1203.

Garí, E., Piedrafita, L., Aldea, M., & Herrero, E. (1997). A set of vectors with a tetracycline-regulatable promoter system for modulated gene expression in Saccharomyces cervisiae. *Yeast*, *13*, 837–848.

Gentz, R., & Bujard, H. (1985). Promoters recognized by Escherichia coli RNA polymerase selected by function : highly Promoters Recognized by Escherichia coli RNA Polymerase Selected by Function : Highly Efficient Promoters from, *164*.

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*, 343–345.

Giniger, E., Varnum, S. M., & Ptashne, M. (1985). Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell*.

Grkovic, S., Brown, M. H., Roberts, N. J., Paulsen, I. T., & Skurray, R. a. (1998). QacR is a repressor protein that regulates expression of the Staphylococcus aureus multidrug efflux pump QacA. *The Journal of Biological Chemistry*, *273*, 18665–73.

Grkovic, S., Hardie, K. M., Brown, M. H., & Skurray, R. a. (2003). Interactions of the QacR multidrug-binding protein with structurally diverse ligands: implications for the evolution of the binding pocket. *Biochemistry*, *42*, 15226–36.

Grohmann, K., Torget, R., Himmel, M., & Scott, C. D. (ed. . (1986). Dilute acid pretreatment of biomass at high acid concentrations. *Biotechnol. Bioeng. Symp.; (United States)*, *17*.

Guzman, L. M., Belin, D., Carson, M. J., & Beckwith, J. (1995). Tight regulation, modulation, and high-level expression by vectors containing the arabinose P(BAD) promoter. *Journal of Bacteriology*, *177*, 4121–4130.

Handelsman, J., Rondon, M. R., Brady, S. F., Clardy, J., & Goodman, R. M. (1998). Molecular biological access to the chemistry of unknown soil microbes: a new frontier for natural products. *Chemistry & Biology*, *5*, R245–R249.

Henderson, P. J., & Maiden, M. C. (1990). Homologous sugar transport proteins in Escherichia coli and their relatives in both prokaryotes and eukaryotes. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *326*, 391–410.

Higgins, C. F. (2007). Multiple molecular mechanisms for multidrug resistance transporters. *Nature*, *446*, 749–757.

Hillen, W., & Berens, C. (1994). Mechanisms underlying expression of Tn10 encoded tetracycline resistance. *Annual Review of Microbiology*, *48*, 345–369.

Hillen, W., Gatz, C., Altschmied, L., Schollmeier, K., & Meier, I. (1983). Control of expression of the Tn10-encoded tetracycline resistance genes. Equilibrium and kinetic investigation of the regulatory reactions. *Journal of Molecular Biology*, *169*, 707–21.

Hottiger, T., Fürst, P., Pohlig, G., & Heim, J. (1994). Physiological characterization of the yeast metallothionein (CUP1) promoter, and consequences of overexpressing its transcriptional activator, ACE1. *Yeast*, *10*, 283–296.

Huang, Q., Wang, Q., Gong, Z., Jin, G., Shen, H., Xiao, S., Xie, H., Ye, S., … Zhao, Z. K. (2013). Effects of selected ionic liquids on lipid production by the oleaginous yeast Rhodosporidium toruloides. *Bioresource Technology*, *130*, 339–344.

Keasling, J. D. (2008). Synthetic biology for synthetic chemistry. *ACS Chemical Biology*.

Keasling, J. D. (2012). Synthetic biology and the development of tools for metabolic engineering. *Metabolic Engineering*, *14*, 189–195.

Kennell, D., & Riezman, H. (1977). Transcription and translation initiation frequencies of the Escherichia coli lac operon. *Journal of Molecular Biology*, *114*, 1–21.

Keskin, S., Kayraktalay, D., Akman, U., & Hortacsu, O. (2007). A review of ionic liquids towards supercritical fluid applications. *The Journal of Supercritical Fluids*, *43*, 150–180.

Khatri, P., Sirota, M., & Butte, A. J. (2012). Ten years of pathway analysis: current approaches and outstanding challenges. *PLoS Computational Biology*, *8*, e1002375.

Khudyakov, J. I., D'haeseleer, P., Borglin, S. E., DeAngelis, K. M., Woo, H., Lindquist, E. A., Hazen, T. C., Simmons, B. A., & Thelen, M. P. (2012). Global transcriptome response to ionic liquid by a tropical rain forest soil bacterium, Enterobacter lignolyticus. *Proc. Natl. Acad. Sci.*, *109*, E2173–E2182.

Klein-Marcuschamer, D., Simmons, B. A., & Blanch, H. W. (2011). Techno-economic analysis of a lignocellulosic ethanol biorefinery with ionic liquid pre-treatment. *Biofuels Bioproducts & Biorefining-Biofpr*, *5*, 562–569.

Klinke, H. B., Thomsen, A. B., & Ahring, B. K. (2004). Inhibition of ethanol-producing yeast and bacteria by degradation products produced during pre-treatment of biomass. *Applied Microbiology and Biotechnology*, *66*, 10–26.

Krulwich, T. a, Lewinson, O., Padan, E., & Bibi, E. (2005). Do physiological roles foster persistence of drug/multidrug-efflux transporters? A case study. *Nature Reviews. Microbiology*, *3*, 566–572.

Lamphier, M. S., & Ptashne, M. (1992). Multiple mechanisms mediate glucose repression of the yeast GAL1 gene. *Proc. Natl. Acad. Sci.*, *89*, 5922–5926.

Lee, S. K., Chou, H., Ham, T. S., Lee, T. S., & Keasling, J. D. (2008). Metabolic engineering of microorganisms for biofuels production: from bugs to synthetic biology to fuels. *Current Opinion in Biotechnology*, *19*, 556–563.

Lee□□, S.-M., & Chang, W.-J. (2005). Influence of Ionic Liquids on the Growth of Escherichia coli. *Korean J. Chem. Eng*, *22*, 687–690.

Lee, T. S., Krupa, R. A., Zhang, F., Hajimorad, M., Holtz, W. J., Prasad, N., Lee, S. K., & Keasling, J. D. (2011). BglBrick vectors and datasheets: A synthetic biology platform for gene expression. *J Biol Eng*, *5*, 12.

Lewinson, O., Padan, E., & Bibi, E. (2004). Alkalitolerance: a biological function for a multidrug transporter in pH homeostasis. *Proc. Natl. Acad. Sci.*, *101*, 14073–14078.

Li, C., Cheng, G., Balan, V., Kent, M. S., Ong, M., Chundawat, S. P. S., Sousa, L. daCosta, Melnichenko, Y. B., … Singh, S. (2011). Influence of physico-chemical changes on enzymatic digestibility of ionic liquid and AFEX pretreated corn stover. *Bioresource Technology*, *102*, 6928–6936.

Li, C. L., Knierim, B., Manisseri, C., Arora, R., Scheller, H. V, Auer, M., Vogel, K. P., Simmons, B. A., & Singh, S. (2010). Comparison of dilute acid and ionic liquid pretreatment of switchgrass: Biomass recalcitrance, delignification and enzymatic saccharification. *Bioresource Technology*, *101*, 4900–4906.

Linger, J. G., Vardon, D. R., Guarnieri, M. T., Karp, E. M., Hunsinger, G. B., Franden, M. a., Johnson, C. W., Chupka, G., … Beckham, G. T. (2014). Lignin valorization through integrated biological funneling and chemical catalysis. *Proc. Natl. Acad. Sci.*, *111*, 1410657111–.

Lisser, S., & Margalit, H. (1993). Compilation of E.coli mRNA promoter sequences. *Nucleic Acids Research*, *21*, 1507–1516.

Liu, T. G., & Khosla, C. (2010). Genetic Engineering of Escherichia coli for Biofuel Production. *Annual Review of Genetics, Vol 44*, *44*, 53–69.

Loeschcke, A., Markert, A., Wilhelm, S., Wirtz, A., Rosenau, F., Jaeger, K. E., & Drepper, T. (2013). TREX: A universal tool for the transfer and expression of biosynthetic pathways in bacteria. *ACS Synthetic Biology*, *2*, 22–33.

Lovell, S., Marquardt, B. J., & Kahr, B. (1999). Crystal violet ' s shoulder. *J. Chem. Soc., Perkin Trans. 2*, 2241–2247.

Łuczak, J., Jungnickel, C., Łącka, I., Stolte, S., & Hupka, J. (2010). Antimicrobial and surface activity of 1-alkyl-3-methylimidazolium derivatives. *Green Chemistry*, *12*, 593.

Lussier, F. X., Denis, F., & Shareck, F. (2010). Adaptation of the highly productive T7 expression system to Streptomyces lividans. *Applied and Environmental Microbiology*, *76*, 967–970.

Martin, V. J. J., Pitera, D. J., Withers, S. T., Newman, J. D., & Keasling, J. D. (2003). Engineering a mevalonate pathway in Escherichia coli for production of terpenoids. *Nature Biotechnology*, *21*, 796–802.

Marx, C. J., & Lidstrom, M. E. (2001). Development of improved versatile broad-host-range vectors for use in methylotrophs and other Gram-negative bacteria. *Microbiology (Reading, England)*, *147*, 2065–75.

Maya, D., Quintero, M. J., de la Cruz Munoz-Centeno, M., & Chavez, S. (2008). Systems for applied gene control in Saccharomyces cerevisiae. *Biotechnol Lett*, *30*, 979–987.

Megaw, J., Busetti, A., & Gilmore, B. F. (2013). Isolation and Characterisation of 1-Alkyl-3-Methylimidazolium Chloride Ionic Liquid-Tolerant and Biodegrading Marine Bacteria. *PLoS ONE*, *8*, 1–9.

Mora-Pale, M., Meli, L., Doherty, T. V, Linhardt, R. J., & Dordick, J. S. (2011). Room Temperature Ionic Liquids as Emerging Solvents for the Pretreatment of Lignocellulosic Biomass. *Biotechnology and Bioengineering*, *108*, 1229–1245.

Murakami, S., Tamura, N., Saito, A., Hirata, T., & Yamaguchi, A. (2004). Extramembrane Central Pore of Multidrug Exporter AcrB in Escherichia coli Plays an Important Role in Drug Transport. *Journal of Biological Chemistry*, *279*, 3743–3748.

Mutalik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A., Tran, A. B., Paull, M., … Endy, D. (2013). Precise and reliable gene expression via standard transcription and translation initiation elements. *Nature Methods*, *10*, 354–60.

Nagashima, K., Kasai, M., Nagata, S., & Kaziro, Y. (1986). Structure of the two genes coding for polypeptide chain elongation factor 1 alpha (EF-1 alpha) from Saccharomyces cerevisiae. *Gene*, *45*, 265–73.

Neufeld, J. D., Engel, K., Cheng, J., Moreno-Hagelsieb, G., Rose, D. R., & Charles, T. C. (2011). Open resource metagenomics: a model for sharing metagenomic libraries. *Stand Genomic Sci*, *5*, 203–210.

Nevoigt, E., Kohnke, J., Fischer, C. R., Alper, H., Stahl, U., & Stephanopoulos, G. (2006). Engineering of promoter replacement cassettes for fine-tuning of gene expression in Saccharomyces cerevisiae. *Applied and Environmental Microbiology*, *72*, 5266–5273.

Nevozhay, D., Adams, R. M., Murphy, K. F., Josic, K., & Balázsi, G. (2009). Negative autoregulation linearizes the dose-response and suppresses the heterogeneity of gene expression. *Proc. Natl. Acad. Sci.*, *106*, 5123–5128.

Neyfakh, A. a. (2002). Mystery of multidrug transporters: The answer can be simple. *Molecular Microbiology*, *44*, 1123–1130.

Nikaido, H. (2009). Multidrug resistance in bacteria. *Annual Review of Biochemistry*, *78*, 119–146.

Ogawa, W., Koterasawa, M., Kuroda, T., & Tsuchiya, T. (2006). KmrA multidrug efflux pump from Klebsiella pneumoniae. *Biological & Pharmaceutical Bulletin*, *29*, 550–553.

Ouellet, M., Datta, S., Dibble, D. C., Tamrakar, P. R., Benke, P. I., Li, C. L., Singh, S., Sale, K. L., … Mukhopadhyay, A. (2011). Impact of ionic liquid pretreated plant biomass on Saccharomyces cerevisiae growth and biofuel production. *Green Chemistry*, *13*, 2743–2749.

Pao, S. S., Paulsen, I. A. N. T., & Saier, M. H. (1998). Major Facilitator Superfamily. *Microbiology and Molecular Biology Reviews*, *62*, 1–34.

Park, J. I., Steen, E. J., Burd, H., Evans, S. S., Redding-Johnson, A. M., Batth, T., Benke, P. I., D'haeseleer, P., … Gladden, J. M. (2012). A Thermophilic Ionic Liquid-Tolerant Cellulase Cocktail for the Production of Cellulosic Biofuels. *Plos One*, *7*.

Partow, S., Siewers, V., Bjørn, S., Nielsen, J., & Maury, J. (2010). Characterization of different promoters for designing a new expression vector in Saccharomyces cerevisiae. *Yeast*, *27*, 955–964.

Paulsen, I. T., Brown, M. H., & Skurray, R. A. (1996). Proton-dependent multidrug efflux systems. *Microbiological Reviews*, *60*, 575–608.

Pedelacq, J. D., Cabantous, S., Tran, T., Terwilliger, T. C., & Waldo, G. S. (2006). Engineering and characterization of a superfolder green fluorescent protein (vol 24, pg 79, 2005). *Nature Biotechnology*, *24*, 1170.

Peralta-Yahya, P. P., Ouellet, M., Chan, R., Mukhopadhyay, A., Keasling, J. D., & Lee, T. S. (2011). Identification and microbial production of a terpene-based advanced biofuel. *Nature Communications*, *2*, 483.

Peters, K. M., Brooks, B. E., Schumacher, M. A., Skurray, R. a, Brennan, R. G., & Brown, M. H. (2011). A single acidic residue can guide binding site selection but does not govern QacR cationic-drug affinity. *PloS One*, *6*, e15974.

Pfeifer, B. A., Admiraal, S. J., Gramajo, H., Cane, D. E., & Khosla, C. (2001). Biosynthesis of complex polyketides in a metabolically engineered strain of E. coli. *Science (New York, N.Y.)*, *291*, 1790–2.

Quan, J. Y., & Tian, J. D. (2009). Circular Polymerase Extension Cloning of Complex Gene Libraries and Pathways. *Plos One*, *4*.

Rabinovitch-Deere, C. A., Oliver, J. W., Rodriguez, G. M., & Atsumi, S. (2013). Synthetic biology and metabolic engineering approaches to produce biofuels. *Chem Rev*, *113*, 4611–4632.

Ramos, J. L., Martı, M., Molina-henares, A. J., Tera, W., Brennan, R., & Tobes, R. (2005). The TetR Family of Transcriptional Repressors. *Microbiology and Molecular Biology Reviews*, *69*, 326–356.

Ranke, J., Müller, A., Bottin-Weber, U., Stock, F., Stolte, S., Arning, J., Störmann, R., & Jastorff, B. (2007). Lipophilicity parameters for ionic liquid cations and their correlation to in vitro cytotoxicity. *Ecotoxicology and Environmental Safety*, *67*, 430–438.

Ro, D.-K., Paradise, E. M., Ouellet, M., Fisher, K. J., Newman, K. L., Ndungu, J. M., Ho, K. A., Eachus, R. A., … Keasling, J. D. (2006). Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature*, *440*, 940–943.

Ruegg, T. L., Kim, E.-M., Simmons, B. a, Keasling, J. D., Singer, S. W., Soon Lee, T., & Thelen, M. P. (2014). An auto-inducible mechanism for ionic liquid resistance in microbial biofuel production. *Nature Communications*, *5*, 3490.

Saier, M. H., & Paulsen, I. T. (2001). Phylogeny of multidrug transporters. *Seminars in Cell & Developmental Biology*, *12*, 205–213.

Saier, M. H., Paulsen, I. T., Sliwinski, M. K., Pao, S. S., Skurray, R. a, & Nikaido, H. (1998). Evolutionary origins of multidrug and drug-specific efflux pumps in bacteria. *The FASEB Journal : Official Publication of the Federation of American Societies for Experimental Biology*, *12*, 265–274.

Samorì, C. (2011). Ionic Liquids and their Biological Effects Towards Microorganisms. *Current Organic Chemistry*, 1888–1904.

Santiviago, C. A., Fuentes, J. A., Bueno, S. M., Trombert, A. N., Hildago, A. A., Socias, L. T., Youderian, P., & Mora, G. C. (2002). The Salmonella enterica sv. Typhimurium smvA, yddG and ompD (porin) genes are required for the efficient efflux of methyl viologen. *Molecular Microbiology*, *46*, 687–698.

Schumacher, M. A., Miller, M. C., & Brennan, R. G. (2004). Structural mechanism of the simultaneous binding of two drugs to a multidrug-binding protein. *The EMBO Journal*, *23*, 2923–2930.

Schumacher, M. A., Miller, M. C., Grkovic, S., Brown, M. H., Skurray, R. a, & Brennan, R. G. (2001). Structural mechanisms of QacR induction and multidrug recognition. *Science*, *294*, 2158–2163.

Shank, E. a., Klepac-Ceraj, V., Collado-Torres, L., Powers, G. E., Losick, R., & Kolter, R. (2011). PNAS Plus: Interspecies interactions that result in Bacillus subtilis forming biofilms are mediated mainly by members of its own genus. *Proceedings of the National Academy of Sciences*, *108*, E1236–E1243.

Sharfstein, S. T., Van Dien, S. J., & Keasling, J. D. (1996). Modulation of the phosphate-starvation response in Escherichia coli by genetic manipulation of the polyphosphate pathways. *Biotechnology and Bioengineering*, *51*, 434–435.

Shen, A., Lupardus, P. J., Morell, M., Ponder, E. L., Sadaghiani, A. M., Garcia, K. C., & Bogyo, M. (2009). Simplified, enhanced protein purification using an inducible, autoprocessing enzyme tag. *PLoS ONE*, *4*.

Shimada, T., Yamazaki, Y., Tanaka, K., & Ishihama, A. (2014). The whole set of constitutive promoters recognized by RNA polymerase RpoD holoenzyme of Escherichia coli. *PLoS ONE*, *9*.

Sikorski, R. S., & Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in Saccharomyces cerevisiae. *Genetics*, *122*, 19–27.

Sitepu, I. R., Shi, S., Simmons, B. A., Singer, S. W., Boundy-mills, K., & Simmons, C. W. (2014). Yeast tolerance to the ionic liquid 1-ethyl-3-methylimidazolium acetate. *FEMS Yeast Research*, *14*, 1286–1294.

Skerra, A. (1994). Use of the tetracycline promoter for the tightly regulated production of a murine antibody fragment in Escherichia coli. *Gene*, *151*, 131–135.

Socha, A. M., Parthasarathi, R., Shi, J., Pattathil, S., Whyte, D., Bergeron, M., George, A., Tran, K., … Singh, S. (2014). Efficient biomass pretreatment using ionic liquids derived from lignin and hemicellulose. *Proc. Natl. Acad. Sci.*, *111*, E3587–E3595.

Sørensen, H. P., & Mortensen, K. K. (2005). Advanced genetic strategies for recombinant protein expression in Escherichia coli. *Journal of Biotechnology*, *115*, 113–128.

Stanton, B. C., Nielsen, A. a K., Tamsir, A., Clancy, K., Peterson, T., & Voigt, C. a. (2014). Genomic mining of prokaryotic repressors for orthogonal logic gates. *Nature Chemical Biology*, *10*, 99–105.

Steen, E. J., Kang, Y., Bokinsky, G., Hu, Z., Schirmer, A., McClure, A., Del Cardayre, S. B., & Keasling, J. D. (2010). Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature*, *463*, 559–62.

Stepnowski, P., Müller, A., Behrend, P., Ranke, J., Hoffmann, J., & Jastorff, B. (2003). Reversed-phase liquid chromatographic method for the determination of selected room-temperature ionic liquid cations. *Journal of Chromatography A*, *993*, 173–178.

Tabor, S., & Richardson, C. C. (1985). A bacteriophage T7 RNA polymerase/promoter system for controlled exclusive expression of specific genes. *Proc. Natl. Acad. Sci.*, *82*, 1074–1078.

Tahlan, K., Ahn, S. K., Sing, A., Bodnaruk, T. D., Willems, A. R., Davidson, A. R., & Nodwell, J. R. (2007). Initiation of actinorhodin export in Streptomyces coelicolor. *Molecular Microbiology*, *63*, 951–961.

Terpe, K. (2006). Overview of bacterial expression systems for heterologous protein production: From molecular and biochemical fundamentals to commercial systems. *Applied Microbiology and Biotechnology*.

Terrón-González, L., Medina, C., Limón-Mortés, M. C., & Santero, E. (2013). Heterologous viral expression systems in fosmid vectors increase the functional analysis potential of metagenomic libraries. *Scientific Reports*, *3*, 1107.

Troeschel, S. C., Drepper, T., Leggewie, C., Streit, W. R., & Jaeger, K. (2010). Novel Tools for the Functional Expression of Metagenomic DNA. *Metagenomics Methods and Protocols*, *668*, 117–139.

Uchiyama, T., Abe, T., Ikemura, T., & Watanabe, K. (2005). Substrate-induced gene-expression screening of environmental metagenome libraries for isolation of catabolic genes. *Nature Biotechnology*, *23*, 88–93.

Ulrich, L. E., Koonin, E. V., & Zhulin, I. B. (2005). One-component systems dominate signal transduction in prokaryotes. *Trends in Microbiology*.

Wade, H. (2010). MD recognition by MDR gene regulators. *Current Opinion in Structural Biology*, *20*, 489–496.

Wagner, S., Baars, L., Ytterberg, A. J., Klussmeier, A., Wagner, C. S., Nord, O., Nygren, P. A., van Wijk, K. J., & de Gier, J. W. (2007). Consequences of membrane protein overexpression in Escherichia coli. *Molecular & Cellular Proteomics*, *6*, 1527–1550.

Wang, Y., Mukhopadhyay, A., Howitz, V. R., Binns, A. N., & Lynn, D. G. (2000). Construction of an efficient expression system for Agrobacterium tumefaciens based on the coliphage T5 promoter. *Gene*, *242*, 105–114.

Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., & Barton, G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics*, *25*, 1189–1191.

Weber, T., Welzel, K., Pelzer, S., Vente, A., & Wohlleben, W. (2003). Exploiting the genetic potential of polyketide producing streptomycetes. *Journal of Biotechnology*, *106*, 221–32.

Wen, M., Bond-Watts, B. B., & Chang, M. C. (2013). Production of advanced biofuels in engineered E. coli. *Curr Opin Chem Biol*, *17*, 472–479.

Woo, H. L., Ballor, N. R., Hazen, T. C., Fortney, J. L., Simmons, B., Davenport, K., Goodwin, L., Ivanova, N., … DeAngelis, K. M. (2014). Complete genome sequence of the lignin-degrading bacterium Klebsiella sp. strain BRL6-2. *Standards in Genomic Sciences*, *9*, 19.

Wu, K., & Rao, C. V. (2010). The role of configuration and coupling in autoregulatory gene circuits. *Molecular Microbiology*, *75*, 513–527.

Yoon, H., Ansong, C., McDermott, J. E., Gritsenko, M., Smith, R. D., Heffron, F., & Adkins, J. N. (2011). Systems analysis of multiple regulator perturbations allows discovery of virulence factors in Salmonella. *BMC Systems Biology*, *5*, 100.

Zheleznova, E. E., Markham, P. N., Neyfakh, A. a., & Brennan, R. G. (1999). Structural basis of multidrug recognition by BMRR, a transcription activator of a multidrug transporter. *Cell*, *96*, 353–362.

ACKNOWLEDGMENTS

First of all, I would like to sincerely thank Prof. Thomas Boller and Dr. Michael Thelen for making my project become true. I am very grateful to Thomas for letting me pursue my research ideas in quite distant fields and continents. I received Thomas' immediate support during my Master's thesis in Panamá and, several years later, for this PhD project. At this point, I also thank Prof. Martin Ackermann who willingly accepted to join the thesis committee.

I am deeply thankful to Michael for receiving his support throughout my research in every sense. He always had two open ears for my questions and gave me lots of freedom to explore, experiment and collaborate. It has been a big pleasure to work with Michael, with whom I enjoyed great moments inside and outside the lab. I thank Dr. Blake Simmons for supporting my research ideas and for never hesitating to provide funding for travels, whether it involved sampling in Puerto Rico, attending conferences or an extensive visit to Prof. Jo Handelsman's lab. At this point, I would like to thank Jo for introducing me into the world of functional metagenomics and for inspiring discussions.

I enjoyed the inspiring chats with Jim Kirby, Greg Linshiz, Vivek Mutalik, Greg Bokinsky, Dominique Loque, Gabriella Papa and Doug Higgins, some of which triggered new ideas. I would especially like to thank Adrienne McKee, with whom I enjoyed many stimulating discussions and who was very patient while introducing me into the world of molecular biology. It was great to work in numerous collaborations with Pavel Novichkov, Henrique Pereira, Joseph Chen, Henning Kirst, Paloma Rueda, Eunmi Kim, and Erika Yoshida. I could learn enormously beyond my immediate research. I also appreciate the useful practical help and advice from Andy DeGiovanni, Richard Heins, Keefe Holland, and Hannah Woo. Many of these persons became good friends.

I also would like to thank Aindrila Mukhopadhyay, Dianelys Saens and David Serp for establishing a parallel connection line to JBEI.

I am very happy that Leticia and I shared the curiosity of moving to California. Thank you very much for always being supportive throughout my research project and for understanding my occasional absentmindedness in the final stage.

I thank my family for their support, specifically my father for proofreading my thesis and my brother for designing the cover page.

I am very grateful to the Swiss National Science Foundation, whose support was essential for the development of promoters. I would also like to thank the Emilia Guggenheim-Schnurr Foundation and the Freiwillige Akademische Gesellschaft Basel for additional financial support.

List of Abbreviations

bp	basepair
[C_2mim]Cl	1-ethyl-3-methylimidazolium chloride
colE1	origin of replication (high copy number)
CPEC	Circular polymerase extension cloning
EMSA	Electrophoretic mobility shift assay
EZ	EZ-Rich defined medium containing 0.2% glucose (unless otherwise stated)
FACS	Fluorescence-activated cell sorting
IL	Ionic liqud
IPTG	Isopropyl β-D-1-thiogalactopyranoside
LB	Lysogeny broth
MDR	Multidrug resistance
MFS	Major facilitator superfamily
p15A	origin of replication (medium copy number)
PCR	Polymerase chain reaction
PYE	Peptone yeast extract
RFP	Red fluorescent protein
RFU	Relative fluorescence units (here: fluorescence signal divided by cell density)
SC101	origin of replication (low copy number)
TAE	Tris-acetate-EDTA buffer
TB	Terrific broth, here containing 2% glycerol
YFP	Yellow fluorescent protein